511,455

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
23 October 2003 (23.10.2003)

PCT

(10) International Publication Number
WO 03/087408 A2

(51) International Patent Classification[7]: C12Q 1/68, A61K 38/00, 48/00, 39/395, A01K 67/027, G01N 33/53

(21) International Application Number: PCT/GB03/01543

(22) International Filing Date: 7 April 2003 (07.04.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

GB 0207902.8	5 April 2002 (05.04.2002)	GB
GB 0207904.4	5 April 2002 (05.04.2002)	GB
GB 0207900.2	5 April 2002 (05.04.2002)	GB
GB 0207901.0	5 April 2002 (05.04.2002)	GB
GB 0227734.1	28 November 2002 (28.11.2002)	GB
GB 0227733.3	28 November 2002 (28.11.2002)	GB

(71) Applicant *(for all designated States except US)*: **UNIVERSITY COURT OF THE UNIVERSITY OF EDINBURGH** [GB/GB]; Old College, South Bridge, Edinburgh EH8 9YL (GB).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **PICKARD, Benjamin, Simon** [GB/GB]; 2 Noble Place, Edinburgh EH6 8AX (GB). **BLACKWOOD, Douglas** [GB/GB]; 7 Pitsligo Road, Edinburgh EH10 4RY (GB). **PORTEOUS, David, John** [GB/GB]; Cherrytrees, 8 Cammo Gardens, Edinburgh EH4 8EH (GB). **MUIR, Walter, John** [GB/GB]; Chestnuts, Eshiels, Peebles, Tweeddale EH45 8NA (GB). **MORS, Ole** [DK/DK]; Permelillevej 7, DK-8240 Risskov (DK). **EWALD, Henrik, Lykke** [DK/DK]; Montanagade 9, DK-8000 Aarhus C (DK).

(74) Agents: **MACDOUGALL, Donald, Carmichael** et al.; Cruikshank & Fairweather, 19 Royal Exchange Square, Glasgow G1 3AE (GB).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 03/087408 A2

(54) Title: SCHIZOPHRENIA ASSOCIATED GENES

(57) **Abstract:** The present invention relates to the identification of genes which have been disrupted in patients diagnosed as suffering from schizophrenia and/or bi-polar affective disorder, as well as proteins encoded by the gene and antibodies thereto and to uses of such products as medicaments for treating schizophrenia and/or affective psychosis. The invention also relates to methods for diagnosing patients suffering or predisposed to schizophrenia and/or affective psychosis, as well as screens for developing novel treatment regimes for schizophrenia and/or affective psychosis.

SCHIZOPHRENIA ASSOCIATED GENES

The present invention relates to the identification of genes which have been disrupted in patients diagnosed as suffering from schizophrenia and/or bi-polar affective disorder, as well as proteins encoded by the gene and antibodies thereto and to uses of such products as medicaments for treating schizophrenia and/or affective psychosis. The invention also relates to methods for diagnosing patients suffering or predisposed to schizophrenia and/or affective psychosis, as well as screens for developing novel treatment regimes for schizophrenia and/or affective psychosis.

Schizophrenia and Bipolar Affective Disorder are common and debilitating psychiatric disorders. Despite a wealth of information on the epidemiology, neuroanatomy and pharmacology of the illness, it is uncertain what molecular pathways are involved and how impairments in these affect brain development and neuronal function. Despite an estimated heritability of 60-80%, very little is known about the number or identity of genes involved in these psychoses. Although there has been recent progress in linkage and association studies, especially from genome-wide scans, these studies have yet to progress from the identification of susceptibility loci or candidate genes to the full characterisation of disease-causing genes (Berrettini, 2000).

The cloning of breakpoints in patients with chromosome abnormalities (translocations, inversions etc.) has proved instrumental in the identification of many disease genes (e.g. Duchenne Muscular Dystrophy, Retinoblastoma, Wilm's Tumour, Familial Polyposis Coli, Fragile-X Syndrome, Polycystic Kidney Disease, many leukaemias and, very recently, a candidate speech and language disorder gene (Lai et al, 2001)). Such studies assume that the chromosomal breakpoints give rise to the clinical symptoms by either directly disrupting gene sequences or perturbing gene expression. In the same way that gene-trap

mutagenesis can be used to identify disrupted mouse genes (Brennan & Skarnes, 1999), the physical "flag" created by a cytogenetic breakpoint provides a geographical pointer for the disease locus.

It is amongst the objects of the present invention to provide genes and/or proteins postulated to be involved with the development and/or symptoms associated with schizophrenia and/or affective psychosis.

As will be seen, the present invention is based on the molecular characterisation of a chromosomal disruption in subjects diagnosed as suffering from a schizophrenia and/or affective psychosis. A high-throughput Fluorescence *in situ* Hybridisation (FISH)-based approach has been adopted to map the chromosomal breakpoints in these patients. Consultation of the sequence data at the breakpoint locus not only allows efficient FISH probe selections to be made by the targeting of coding regions, but also proof of gene disruption can be made entirely by relating the exact position of probes to the genomic structure of a candidate gene.

Four patients have been studied and their chromosomal disruptions characterised. Hereinafter the patients will be identified as patients 1-4.

As will be seen, in one embodiment, the present invention is based on the molecular characterisation of a chromosomal rearrangement denoted t(3;8)(p13;p22) in a subject (patient 1) diagnosed as suffering from a schizoaffective disorder (see Fig.1). A high-throughput Fluorescence *in situ* Hybridisation (FISH)-based approach was adopted to map the chromosomal breakpoints in these patients. Consultation of the sequence data at the breakpoint loci not only allowed efficient FISH probe selections to be made by the targeting of coding regions, but also proof of gene disruption was inferred entirely by relating the exact position of probes to the genomic structure of a candidate gene.

One breakpoint (located on chromosome 8p22) in this subject lies near to a gene, *N33,* involved in the N-Linked Glycosylation pathway.

This pathway consists of three stages. Firstly the assembly of a donor oligosaccharide at the endoplasmic reticulum lumen membrane. Secondly, the transfer of this molecule onto newly translated secretory and transmembrane proteins catalyzed by the oligosaccharyltransferase (OST) complex. Thirdly, there is subsequent modification of the oligosaccharides on the glycoprotein. N33 encodes a protein thought to be involved in the second stage of the pathway by analogy with yeast homologues. Without wishing to be bound by theory it is hypothesised that the breakpoint in the subject perturbs N33 expression indirectly through position effect silencing or separation of regulatory elements from the gene promoter (both effects have been shown to occur even when the breakpoints are up to 1Mb from the target gene in some instances (Kleinjan et al 1999)).

As the *N33* gene is located within a chromosomal region repeatedly found positive in schizophrenia linkage studies the present inventors pursued this gene further by association study.

Certain microsatellite repeat haplotypes have been identified at the N33 locus which are over-represented in schizophrenic patients and their families compared to the normal population. Subsequent sequencing of the *N33* gene in haplotype carrying individuals is ongoing in order to identify causative mutations.

The other breakpoint in this patient (3p13) has now also been fully characterised and demonstrated to disrupt a gene, *SEMCAP3* (also known as KIAA1095). The present invention is therefore also based on a proposed role of this gene (normal and mutated forms) in the aetiology of schizophrenia and/or affective psychosis.

In a further embodiment the present invention is based on the *GRIK4* gene and observations of the present inventors of an involvement of this gene and/or protein with schizophrenia and/or affective psychosis.

The *GRIK4* gene is also known as KA1 and EAA1, but will herein be referred to as *GRIK4* for simplicity, but should not be construed as limiting.

The subject (patient 2) was one of a series of around 100 patients with comorbid schizophrenia and mild learning disability (US terminology: "mental retardation") who were screened using routine G-band karyotyping. This patient possesses a complex chromosomal rearrangement which can be described by standard nomenclature as; (46, XX, ins(8;11) (q13;q23.3q24.2)inv(2)(p12q32.1)t(2;11)(q21.3;q24.2)der (2)(2qter->2q32.1::2p12->2q21.3::11q24.2->11qter)der(11)(11pter->11q23.3::2q21.3->2q32.1::2p12->2pter)der(8)(8pter->8q13::11q23.2->11q24.2::8q13->8qter)). It has been repeatedly observed that schizophrenia occurs more frequently in individuals with mild learning disability than in the general population and recent work has revealed an increased heritability of this comorbid state.

As described herein the FISH results reveal that the subject has a disruption in a brain expressed gene; namely, *GRIK4* which is known to participate in molecular mechanisms responsible for modulating the strength of synaptic transmission.

In a further embodiment the present invention is based on the characterisation of a balanced reciprocal translocation between chromosomes 9 and 14, t(9;14)(q34;q13) in a mother (patient 3) with schizophrenia and her daughter with schizophrenia co-morbid with mild learning disability. A brain transcription factor gene, *NPAS3*, is shown to be disrupted by the translocation at 14q13. Without wishing to be bound by theory, the present inventors hypothesis is that the disruption of this gene is responsible for the psychotic symptoms exhibited by the

mother and daughter.

As will be seen, the present invention is also based on the molecular characterisation of a chromosomal rearrangement denoted t(1;16)(p31.2;q21) (in patient 4).

The proband met ICD-10 and DSM-IV criteria for definite schizophrenia. The translocation was inherited within other branches of the family with variable clinical expression. However some key translocation carriers of the subjects to whom the inventors had access had not passed the age of risk when clinically characterized.

One breakpoint (located on chromosome 1p31.2) in patient 4 lies within an alternatively spliced form of the gene, *PDE4B,* involved in the attenuation of cAMP secondary messenger signaling.

The remaining breakpoint in this patient (16q21) has now also fully characterised and demonstrated to disrupt a gene, *CADHERIN 8* (*CDH8*). The present invention is therefore based in part on a proposed role of this gene in the aetiology of schizophrenia and/or affective psychosis.

In a first aspect the present invention provides use of a polynucleotide fragment or fragments comprising *SEMCAP3*, *N33*, *NPAS3*, *GRIK4*, *PDE4B* and/or *CDH8* gene(s) or a fragment(s), derivative(s) or homologue(s) thereof for the manufacture of a medicament for treating schizophrenia and/or affective psychosis in a subject.

In another aspect the present invention provides use of a polypeptide fragment or fragments encoded by *SEMCAP3*, *N33*, *NPAS3*, *GRIK4*, *PDE4B* and/or *CDH8* gene(s), or a fragment(s), derivative(s) or homologue(s) thereof for the manufacture of a medicament for treating schizophrenia and/or affective psychosis in a subject.

Schizophrenia and/or affective psychosis as used herein relates to schizophrenia, as well as other affective psychoses such as those listed in "The ICD-10 Classification of Mental and Behavioural Disorders" World Health Organization, Geneva 1992. Categories F20 to F29 inclusive includes Schizophrenia, schizotypal and

delusional disorders. Categories F30 to F39 inclusive are Mood (affective) disorders that include bipolar affective disorder and depressive disorder. Mental Retardation is coded F70 to F79 inclusive. The Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV). American Psychiatric Association, Washington DC. 1994. Include all conditions coded 295.xx (Schizophrenia and Other Psychotic Disorders) and 296.xx (Major Depressive Disorders and Bipolar Disorders). Mental Retardation is coded 315, 317, 318 and 319.

SEMCAP3 has been previously cloned and sequenced in mouse as two alternative forms (Semcap3*A* and *3B*) and the sequences are present in the public database (nucleic acid sequences; AF127084/AF127085, respectively; protein sequences AAF22131/AAF22132, respectively) as directly submitted by Wang & Strittmatter, 1999. The human form of the gene is defined by sequence KIAA1095 (nucleic acid sequence, AB029018 or XM_041363, and a smaller form, BC014432; protein sequence, XP_041363). The genomic sequences corresponding to this gene are also present in the public database (eg. for BAC RP11-252o10, AC024102). Nevertheless, the prior art does not suggest any link between *SEMCAP3* and schizophrenia and/or affective psychosis.

Thus, references herein to the *SEMCAP3* gene are understood to relate to the sequences in the public databases and identified in Fig.3 and references to the SEMCAP3 protein sequence is understood to relate to the sequences in the public databases and identified in Fig.4.

N33 has been previously cloned and sequenced and the sequence is present in the public database (Nucleic acid sequence; U42349, Protein sequence; Q13454) and described in MacGrogan et al, 1996. The genomic sequences corresponding to this gene are also present in the public database (eg. for BAC RP11-23j14) but some SNP polymorphisms or sequencing errors (eg. an extra "C" present in exon 1b, see hereinafter - cctgccccCaccggg - may

result in differences to the sequences presented herein. Nevertheless, the prior art does not suggest any link between N33 and schizophrenia and affective psychosis.

In addition to the sequences previously identified, the present inventors have identified a new start exon (1a, see Figures 6 and 7) and have observed the complexity of the exon splicing at the 3' end of the gene (see Figures 6 and 7).

Thus, references herein to the *N33* gene are understood to relate to the sequences in the public databases and identified in Figures 6 and 7 and references to the N33 protein sequence are understood to relate to the sequences in the public databases and identified in Figures 6 and 7.

The *GRIK4* gene is located on chromosome 11, at cytogenetic position 11q22.3. The gene encodes a kainate receptor subunit and has been previously described by Kamboj et al, 1994. The cDNA nucleotide sequence and peptide sequence was disclosed by Kamboj et al, 1994 and submitted to the Genbank/EMBL database under accession NM_014619. The coding sequence of the gene is identified as being 2871 nucleotides in length, coding for a protein 957 amino acids. The nucleotide and protein sequences are shown in Figures 10 and 11 respectively. The present inventors have identified an alternative start site for the gene (see Figures 15 - 17) which would result in a shorter gene/protein of 933 amino acids as opposed to 956. The full nucleotide sequence and protein sequence of this alternatively encoded gene/protein is shown in Figures 16 and 17.

Thus, references herein to the *GRIK4* gene are understood to relate to the sequences identified in Figures 10 and 16 and references to the GRIK4 protein sequence are understood to relate to the sequences identified in Figures 11 and 17.

The human form of *NPAS3* has previously been identified and is found in the public database under accession numbers AB054575 and AF164438, with the differences due to

alternative splicing and all forms are encompassed within the present invention.

Thus, references herein to the *NPAS3* gene are understood to relate to the sequences identified in Figures 18 and 20 and references to the NPAS3 protein sequence are understood to relate to the sequences identified in Figures 19 and 21.

The *PDE4B* gene is located on chromosome 1 at cytogenetic position 1p31.2. The gene encodes a phosphodiesterase which shows homology to the *Dunce* leaning and memory gene product of *Drosophila melanogaster*, Bolger et al, 1993. Two long (PDE4B1 and PDE4B3) and one short (PDE4B2) splice form are described herein. There is a core protein sequence of 525 amino acid residues shared by all three forms. On to this is added 39 N-terminal amino acid residues in the case of PDE4B2. Both of the long forms share an additional central stretch of 118 amino acid residues, but then diverge at the N-terminal end of the proteins; PDE4B1 has 93 specific residues and PDE4B3, 78. It is predicted that only the PDE4B1 splice form (brain expressed) may be disrupted by the chromosomal abnormality observed in the patient and family.

Thus, references herein to the *PDE4B* gene are understood to relate to the sequences identified in Figures 25, 27 and 29 and references to the PDE4B protein sequence are understood to relate to the sequences identified in Figures 26, 28 and 30.

CADHERIN 8 (*CDH8*) has been previously cloned and sequenced and the sequence is present in the public database (nucleic acid sequence; L34060/AB035305/NM_001796, protein sequence; NP_001787) and described in Suzuki *et al.*, 1991, Tanihara *et al.*, 1994, and Shimoyama *et al.*, 2000. An alternative transcript form has been described in the rat in which there is a truncation within the 5th cadherin domain (Kido et al., 1998 and see Fig.4). The accession numbers for the normal and truncated forms of *CDH8* in rat are AB010436 and AB010437, respectively. The

corresponding human truncated transcript is not present in the public database and so is not yet confirmed. The genomic sequences corresponding to *CDH8* are also present in the public database (eg. BAC CTC-420A11; AC040161). Nevertheless, the prior art does not suggest any link between *CDH8* and schizophrenia and/or affective psychosis.

Thus, references herein to the *CDH8* gene are understood to relate to the nucleic sequences in the public databases and identified in Fig.35 and references to the CDH8 protein sequences are understood to relate to the sequences in the public databases and identified in Fig.36.

In certain jurisdictions claims to methods of treatment are permissible and so the skilled reader will appreciate that the/said *SEMCAP3*, *N33*, *GRIK4*, *NPAS3*, *PDE4B* and/or *CDH8* gene(s), or fragment(s), derivative(s) or homologue(s) thereof; or SEMCAP3, N33, GRIK4, NPAS3, PDE4B and/or CDH8 protein, or functionally active fragment(s), derivative(s), or homologue(s) thereof, may be administered to an individual as a method of treating an individual with schizophrenia and/or affective psychosis.

"Polynucleotide fragment" as used herein refers to a chain of nucleotides such as deoxyribose nucleic acid (DNA) and transcription products thereof, such as RNA. Naturally, the skilled addressee will appreciate the whole naturally occurring human genome is not included in the definition of polynucleotide fragment.

The polynucleotide fragment can be isolated in the sense that it is substantially free of biological material with which the whole genome is normally associated *in vivo*. The isolated polynucleotide fragment may be cloned to provide a recombinant molecule comprising the polynucleotide fragment. Thus, "polynucleotide fragment includes double and single stranded DNA, RNA and polynucleotide sequences derived therefrom, for example, subsequences of said fragment and which are of any desirable length. Where a nucleic acid is single stranded then both a given strand and a sequence or reverse

complementary thereto is within the scope of the present invention.

In general, the term "expression product" or "gene product" refers to both transcription and translation products of said polynucleotide fragments. When the expression or gene product is a "polypeptide" (i.e. a chain or sequence of amino acids displaying a biological activity substantially similar (eg. 98%, 95%, 90%, 80%, 75% activity) to the biological activity of the protein), it does not refer to a specific length of the product as such. Thus, the skilled addressee will appreciate that "polypeptide" encompasses *inter alia* peptides, polypeptides and proteins. The polypeptide if required, can be modified *in vivo* and *in vitro*, for example by glycosylation, amidation, carboxylation, phosphorylation and/or post-translational cleavage.

The present invention further provides a recombinant or synthetic polypeptide for the manufacture of reagents for use as therapeutic agents in the treatment of schizophrenia and/or affective psychosis. In particular, the invention provides pharmaceutical compositions comprising the recombinant or synthetic polypeptide together with a pharmaceutically acceptable carrier therefor.

The present invention further provides an isolated polynucleotide fragment capable of specifically hybridising to a related polynucleotide sequence from another species. In this manner, the present invention provides probes and/or primers for use in *ex vivo* and/or *in situ* detection and expression studies. Typical detection studies include polymerase chain reaction (PCR) studies, hybridisation studies, or sequencing studies. In principle any specific polynucleotide sequence fragment from the identified sequences may be used in detection and/or expression studies. The skilled addressee understands that a specific fragment is a fragment of the sequence which is of sufficient length, generally greater than 10, 12, 14, 16 or

20 nucleotides in length, to bind specifically to the sequence, under conditions of high stringency, as defined herein, and not bind to unrelated sequences, that is sequences from elsewhere in the genome of the organism other than an allelic form of the sequence or non-homologous sequences from other organisms.

"Capable of specifically hybridising" is taken to mean that said polynucleotide fragment preferably hybridises to a related or similar polynucleotide sequence in preference to unrelated or dissimilar polynucleotide sequences.

The invention includes polynucleotide sequence(s) which are capable of specifically hybridising to an polynucleotide fragment as described herein or to a part thereof without necessarily being completely complementary or reverse complementary to said related polynucleotide sequence or fragment thereof. For example, there may be at least 50%, or at least 75%, at least 90%, or at least 95% complementarity. Of course, in some cases the sequences may be exactly reverse complementary (100% reverse complementary) or nearly so (e.g. there may be less than 10, typically less than 5 mismatches). Thus, the present invention also provides anti-sense or complementary nucleotide sequence(s) which is/are capable of specifically hybridising to the disclosed polynucleotide sequence. If a specific polynucleotide is to be used as a primer in PCR and/or sequencing studies, the polynucleotide must be capable of hybridising to related nucleic acid and capable of initiating chain extension from 3' end of the polynucleotide, but not able to correctly initiate chain extension from unrelated sequences.

If a polynucleotide sequence of the present invention is to be used in hybridisation studies to obtain or identify a related sequence from another organism the polynucleotide sequence should preferably remain hybridised to a sample polynucleotide under stringent conditions. If desired, either the test or sample polynucleotide may be immobilised. Generally the test polynucleotide sequence is

at least 10, 14, 20 or at least 50 bases in length. It may be labelled by suitable techniques known in the art. Preferably the test polynucleotide sequence is at least 200 bases in length and may even be several kilobases in length. Thus, either a denatured sample or test sequence can be first bound to a support. Hybridization can be effected at a temperature of between 50 and 70°C in double strength SSC (2xNaCl 17.5g/l and sodium citrate (SC) at 8.8g/l) buffered saline containing 0.1% sodium dodecyl sulphate (SDS). This can be followed by rinsing of the support at the same temperature but with a buffer having a reduced SSC concentration. Depending upon the degree of stringency required, and thus the degree of similarity of the sequences, such reduced concentration buffers are typically single strength SSC containing 0.1%SDS, half strength SSC containing 0.1%SDS and one tenth strength SSC containing 0.1%SDS. Sequences having the greatest degree of similarity are those the hybridisation of which is least affected by washing in buffers of reduced concentration. It is most preferred that the sample and inventive sequences are so similar that the hybridisation between them is substantially unaffected by washing or incubation in standard sodium citrate (0.1 x SSC) buffer containing 0.1%SDS.

Oligonucleotides may be designed to specifically hybridise to N33 SEMCAP3, NPAS3, GRIK4, PDE4B and/or CDH8 nucleic acid. They may be synthesised, by known techniques and used as primers in PCR or sequencing reactions or as probes in hybridisations designed to detect the presence of a mutated or normal *N33*, *SEMCAP3*, *NPAS3*, *GRIK4*, *PDE4B* and/or *CDH8* gene(s) in a sample. The oligonucleotides may be labelled by suitable labels known in the art, such as, radioactive labels, chemiluminescent labels or fluorescent labels and the like.

The term "oligonucleotide" is not meant to indicate any particular length of sequence and encompasses nucleotides of preferably at least 10b (e.g. 10b to 1kb) in

length, more preferably 12b-500b in length and most preferably 15b to 100b.

The oligonucleotides may be designed with respect to any of the sequences described herein and may be manufactured according to known techniques. They may have substantial sequence identity (e.g. at least 50%, at least 75%, at least 90% or at least 95% sequence identity) with one of the strands shown therein or an RNA equivalent, or with a part of such a strand. Preferably such a part is at least 10, at least 30, at least 50 or at least 200 bases long. It may be an open reading frame (ORF) or a part thereof.

Oligonucleotides which are generally greater than 30 bases in length should preferably remain hybridised to a sample polynucleotide under one or more of the stringent conditions mentioned above. Oligonucleotides which are generally less than 30 bases in length should also preferably remain hybridised to a sample polynucleotide but under different conditions of high stringency. Typically the melting temperature of an oligonucleotide less than 30 bases may be calculated according to the formula of; 2°C for every A or T, plus 4°C for every G or C, minus 5°C. Hybridization may take place at or around the calculated melting temperature for any particular oligonucleotide, in 6 x SSC and 1% SDS. Non specifically hybridised oligonucleotides may then be removed by stringent washing, for example in 3 x SSC and 0.1% SDS at the same temperature. Only substantially similar matched sequences remain hybridised i.e. said oligonucleotide and corresponding test nucleic acid.

When oligonucleotides of generally less than 30 bases in length are used in sequencing and/or PCR studies, the melting temperature may be calculated in the same manner as described above. The oligonucleotide may then be allowed to anneal or hybridise at a temperature around the oligonucleotides calculated melting temperature. In the case of PCR studies the annealing temperature should be

around the lower of the calculated melting temperatures for the two priming oligonucleotides. It is to be appreciated that the conditions and melting temperature calculations are provided by way of example only and are not intended to be limiting. It is possible through the experience of the experimenter to vary the conditions of hybridisation and thus anneal/hybridise oligonucleotides at temperatures above their calculated melting temperature. Indeed this can be desirable in preventing so-called non-specific hybridisation from occurring.

It is possible when conducting PCR studies to predict an expected size or sizes of PCR product(s) obtainable using an appropriate combination of two or more oligonucleotides, based on where they would hybridise to the sequences described herein. If, on conducting such a PCR on a sample of DNA, a fragment of the predicted size is obtained, then this is predictive that the DNA encodes a homologous sequence from a test organism.

Proteins for all the applications described herein can be produced by cloning the gene for example into plasmid vectors that allow high expression in a system of choice e.g. insect cell culture, yeast, animal cells, bacteria such as *Escherichia coli*. To enable effective purification of the protein, a vector may be used that incorporates an epitope tag (or other "sticky" extension such as His6) onto the protein on synthesis. A number of such vectors and purification systems are commercially available.

The polynucleotide fragment can be molecularly cloned into a prokaryotic or eukaryotic expression vector using standard techniques and administered to a host. The expression vector is taken up by cells and the polynucleotide fragment of interest expressed, producing protein.

It will be understood that for the particular polypeptides embraced herein, natural variations such as may occur due to polymorphisms, can exist between individuals or between members of the family. These

variations may be demonstrated by (an) amino acid difference(s) in the overall sequence or by deletions, substitutions, insertions, inversions or additions of (an) amino acid(s) in said sequence. All such derivatives showing the recognised activity are included within the scope of the invention. For example, for the purpose of the present invention conservative replacements may be made between amino acids within the following groups:

(I) Alanine, serine, threonine;

(II) Glutamic acid and aspartic acid;

(III) Arginine and leucine;

(IV) Asparagine and glutamine;

(V) Isoleucine, leucine and valine;

(VI) Phenylalanine, tyrosine and tryptophan

Moreover, recombinant DNA technology may be used to prepare nucleic acid sequences encoding the various derivatives outlined above.

As is well known in the art, the degeneracy of the genetic code permits substitution of bases in a codon resulting in a different codon which is still capable of coding for the same amino acid, e.g. the codon for amino acid glutamic acid is both GAT and GAA. Consequently, it is clear that for the expression of polypeptides from nucleotide sequences described herein or fragments thereof, use can be made of a derivative nucleic acid sequence with such an alternative codon composition different from the nucleic acid sequences shown in the Figures.

The polynucleotide fragments of the present invention are preferably linked to regulatory control sequences. Such control sequences may comprise promoters, operators, inducers, enhancers, silencers, ribosome binding sites, terminators etc. Suitable control sequences for a given host may be selected by those of ordinary skill in the art.

A polynucleotide fragment according to the present invention can be ligated to various expression controlling sequences, resulting in a so called recombinant nucleic acid molecule. Thus, the present invention also includes

an expression vector containing an expressible nucleic acid molecule. The recombinant nucleic acid molecule can then be used for the transformation of a suitable host.

Specific vectors which can be used to clone nucleic acid sequences according to the invention are known in the art (e.g. Rodriguez, R.L. and Denhadt, D.T., Edit., Vectors: a survey of molecular cloning vectors and their uses, Butterworths, 1988, or Jones *et al.*,Vectors: Cloning Applications: Essential Techniques (Essential techniques series), John Wiley & Son. 1998).

The methods to be used for the construction of a recombinant nucleic acid molecule according to the invention are known to those of ordinary skill in the art and are inter alia set forth in Sambrook, et al. (Molecular Cloning: a laboratory manual Cold Spring Harbour Laboratory, 1989).

The present invention also relates to a transformed cell containing the polynucleotide fragment in an expressible form. "Transformation", as used herein, refers to the introduction of a heterologous polynucleotide fragment into a host cell. The method used may be any known in the art, for example, direct uptake, transfection transduction or electroporation (Current Protocols in Molecular Biology, 1995. John Wiley and Sons Inc.). The heterologous polynucleotide fragment may be maintained through autonomous replication or alternatively, may be integrated into the host genome. The recombinant nucleic acid molecules preferably are provided with appropriate control sequences compatible with the designated host which can regulate the expression of the inserted polynucleotide fragment, e.g. tetracycline responsive promoter, thymidine kinase promoter, SV-40 promoter and the like.

Suitable hosts for the expression of recombinant nucleic acid molecules may be prokaryotic or eukaryotic in origin. Hosts suitable for the expression of recombinant nucleic acid molecules may be selected from bacteria, yeast, insect cells and mammalian cells.

In another aspect the present invention also relates to a method of diagnosing schizophrenia and/or affective psychosis or susceptibility to schizophrenia and/or affective psychosis in an individual, wherein the method comprises determining if *SEMCAP3*, *N33*, *GRIK4*, *NPAS3*, *PDE4B* and/or *CDH8* gene(s) in the individual has been disrupted by a mutation or chromosomal rearrangement.

The methods which may be employed to elucidate such a mutation or chromosomal rearrangement are well known to those of skill in the art and could be detected for example using PCR or in hybridisation studies using suitable probes which could be designed to span an identified mutation site or chromosomal breakpoint in close proximity to the/said *N33* SEMCAP3, NPAS3, GRIK4, PDE3B and/or CDH8 gene(s), such as the breakpoint identified by the present inventors and described herein.

Once a particular polymorphism or mutation has been identified it may be possible to determine a particular course of treatment. For example it is known that some forms of treatment work for some patients, but not all. This may in fact be due to mutations in the/said *N33*, *SEMCAP3*, *NPAS3*, *GRIK4*, *PDE4B* and/or *CDH8* gene(s) or surrounding sequence, and it may therefore be possible to determine a treatment strategy using current therapies, based on a patient's genotype.

It will be appreciated that mutations in the gene sequence or controlling elements of a gene, eg. a promoter and/or enhance can have subtle effects such as affecting mRNA splicing/stability/activity and/or control of gene expression levels, which can also be determined. Also the relative levels of RNA can be determined using for example hybridisation or quantitative PCR as a means to determine if the/said *SEMCAP3*, *N33*, *GRIK4*, *NPAS3*, *PDE4B* and/or *CHD8* gene(s) has been disrupted.

Moreover the presence and/or levels of the/said *SEMCAP3*, *N33*, *GRIK4*, *NPAS3*, *PDE4B* and/or *CHD8* gene(s) products themselves can be assayed by immunological

techniques such as radioimmunoassay, Western blotting and ELISA using specific antibodies raised against the gene products. The present invention also therefore relates to antibodies specific for a *SEMCAP3*, *N33*, *GRIK4*, *NPAS3*, *PDE4B* and/or *CHD8* gene(s) product(s) and uses thereof in diagnosis and/or therapy.

A further aspect of the present invention therefore provides antibodies specific to the polypeptides of the present invention or epitopes thereof. Production and purification of antibodies specific to an antigen is a matter of ordinary skill, and the methods to be used are clear to those skilled in the art. The term antibodies can include, but is not limited to polyclonal antibodies, monoclonal antibodies (mAbs), humanised or chimeric antibodies, single chain antibodies, Fab fragments, $F(ab')_2$ fragments, fragments produced by a Fab expression library, anti-idiotypic (anti-Id) antibodies, and epitope binding fragments of any of the above. Such antibodies may be used in modulating the expression or activity of the particular polypeptide, or in detecting said polypeptide *in vivo* or *in vitro*.

Using the sequences disclosed herein, it is possible to identify related sequences in other animals, such as mammals, with the intention of providing an animal model for psychiatric disorders associated with the improper functioning of the nucleotide sequences and proteins of the present invention. Once identified, the homologous sequences can be manipulated in several ways common to the skilled person in order to alter the functionality of the nucleotide sequences and proteins homologous to those of the present invention. For example, "knock-out" animals may be created, that is, the expression of the genes comprising the nucleotide sequences homologous to those of the present invention may be reduced or substantially eliminated in order to determine the effects of reducing or substantially eliminating the expression of such genes. Alternatively, animals may be created where the expression

of the nucleotide sequences and proteins homologous to those of the present invention are upregulated, that is, the expression of the genes comprising the nucleotide sequences homologous to those of the present invention may be increased in order to determine the effects of increasing the expression of these genes. In addition to these manipulations, substitutions, deletions and additions may be made to the nucleotide sequences encoding the proteins homologous to those of the present invention in order to effect changes in the activity of the proteins to help elucidate the function of domains, amino acids, etc. in the proteins. Furthermore, the sequences of the present invention may also be used to transform animals to the manner described above. The manipulations described above may also be used to create an animal model of schizophrenia and/or affective psychosis associated with the improper functioning of the nucleotide sequences and/or proteins of the present invention in order to evaluate potential agents which may be effective for combatting psychotic disorders, such as schizophrenia and/or affective psychosis.

Thus, the present invention also provides for screens for identifying agents suitable for preventing and/or treating schizophrenia and/or affective psychosis associated with disruption or alteration in the expression of the *SEMCAP3*, *N33*, *GRIK4*, *NPAS3*, *PDE3B* and/or *CHD8* gene and/or its gene products. Such screens may easily be adapted to be used for the high throughput screening of libraries of compounds such as synthetic, natural or combinatorial compound libraries.

Thus, the/said *SEMCAP3*, *N33*, *GRIK4*, *NPAS3*, *PDE4B* and/or *CDH8* gene(s) products according to the present invention can be used for the *in vivo* or *in vitro* identification of novel ligands or analogs thereof. For this purpose binding studies can be performed with cells transformed with nucleotide fragments according to the invention or an expression vector comprising a polynucleotide fragment according to the invention, said

cells expressing the/said *SEMCAP3*, *N33*, *GRIK4*, *NPAS3*, *PDE4B* and/or *CDH8* gene(s) products according to the invention.

Alternatively also the/said *SEMCAP3*, *N33*, *GRIK4*, *NPAS3*, *PDE4B* and/or *CDH8* gene(s) products according to the invention as well as ligand-binding domains thereof can be used in an assay for the identification of functional ligands or analogs for the/said *SEMCAP3*, *N33*, *GRIK4*, *NPAS3*, *PDE4B* and/or *CDH8* gene(s) products.

Methods to determine binding to expressed gene products as well as *in vitro* and *in vivo* assays to determine biological activity of gene products are well known. In general, expressed gene product is contacted with the compound to be tested and binding, stimulation or inhibition of a functional response is measured.

Thus, the present invention provides for a method for identifying ligands for *SEMCAP3*, *N33*, *GRIK4*, *NPAS3*, *PDE4B* and/or *CDH8* gene(s) products, said method comprising the steps of:

a) introducing into a suitable host cell a polynucleotide fragment according to the invention;

b) culturing cells under conditions to allow expression of the polynucleotide fragment;

c) optionally isolating the expression product;

d) bringing the expression product (or the host cell from step b)) into contact with potential ligands which will possibly bind to the protein encoded by said polynucleotide fragment from step a);

e) establishing whether a ligand has bound to the expressed protein; and

f) optionally isolating and identifying the ligand.

As a preferred way of detecting the binding of the ligand to the expressed protein, also signal transduction capacity may be measured.

Compounds which activate or inhibit the function of SEMCAP3, *N33*, GRIK4, NPAS3, PDE4B and/or CDH8 gene(s) products may be employed in therapeutic treatments to activate or inhibit the polypeptides of the present

invention.

The present invention will now be further described by way of Example and with reference to the Figures which show:

Figure 1 shows an ideogram diagram of the chromosomal rearrangement (a reciprocal translocation) in patient 1. The two breakpoints are marked at the approximate chromosomal locations at which they are located. In addition, and not to scale, the two candidate disease-causing genes, *N33* and *SEMCAP3*, are placed in the correct orientation and with respect to the breakpoints.

Figure 2 shows a representation of the genomic structure of the *SEMCAP3* gene: its spliced exons spread over a genomic extent of approximately 250kb.
Above the gene, the coding contribution of each exon to the SEMCAP3 protein is indicated by bars and finely dashed lines. The domain structure of SEMCAP3 protein is shown at the top of the figure. 'RING' refers to a RING-finger domain, 'ZF-T.' to a TRAF-type zinc finger (also referred to as a *sina* domain) and 'PDZ' to PDZ domain present in PSD-95, Dlg, and ZO-1/2. The BAC clones used to identify the breakpoint location are included at the bottom of the figure together with the inferred direction (arrows) of the breakpoint from the FISH results using these clones. The heavy dashed line shows the position of the breakpoint with respect to the gene exons and the domain structure of the protein.

Figure 3 Nucleic acid sequence of Human *SEMCAP3* (genomic DNA sequence including CpG island/putative promoter upstream of 5' UTR/cDNA sequence is also included for clarity). The following features are marked for clarity:

a) ATG start site located at position 709 (underlined)

b) GG bases (underlined) at the junction between exons 3 and 4 (i.e. between which the breakpoint is located)

c) UAA stop codon located at position 3907 (underlined).

Figure 4 Amino acid sequence of Human SEMCAP3 with underlined regions of interest.

a) Residues 18-55 Ring finger domain
b) Residues 101-158 SINA/ZF-TRAF domain
c) Residues 246-339 PDZ domain #1
d) Residues 418-504 PDZ domain #2

Figure 5 shows a schematic representation of the *N33* gene : exon splicing and chromosome breakpoint identified in the present invention.

Figure 6 shows the nucleotide sequence of the various exons for N33.

Figure 7 shows the various transcript options and associated amino acid sequences of the transcripts for N33;

Figure 8 shows N33 protein aligned with other homologues.

Figure 9 shows the effect of the C-terminus of the various N33 splice forms. The variety of splice forms at the 3' end of the gene has implications for the C-terminus of the protein. This is especially important when it is considered that N33 is likely to reside in the Golgi/ER compartment of the cell where C-termini are often involved in anchoring or trafficking proteins to different organelles. The light grey shading indicates putative transmembrane domains. Hence, only the spliceforms with exons 1a/1b,2-6,7,8,9,10,11 or 1a/1b,2-6,7,8,9,11 are likely to encode functional proteins and these will only differ in the extreme C-terminal residues.

Figure 10 shows the published nucleotide sequence for GRIK4.

Figure 11 shows the published amino acid sequence for GRIK4.

Figure 12 Breakpoints identified in the subject (patient 2). CEPH library YACs (Chumakov et al, 1992) spanning the breakpoints are listed. Also detailed are the BAC clones (and accession numbers) from the RPCI-11 BAC library (Osoegawa et al, 2001) that span or flank (indicated by dashes) the breakpoints. Breakpoints at 8q13 were not

characterised in this study.

Figure 13 Representation of complex chromosomal rearrangement in the subject (patient 2). The pericentric chromosome 2 inversion is coupled with a translocation to chromosome 11. The chromosome 11 region between the 11q23.3 and 11q24.3 breakpoints is inserted on chromosome 8q13.

Figure 14 Genomic arrangements of the *GRIK4* gene disrupted in the subject. Two potential *GRIK4* transcripts with alternative start-sites are indicated. The 1a/1a' exons are derived from EST BE388730. The transcript

containing the 1b exon corresponds to the published *GRIK4* sequence (acc. S67803). It is probable that the present inventors exon "4" corresponds to a number of undefined exons which can only be subdivided after release of genomic sequence over this part of the gene. Hence, the actual number of *GRIK4* transcript exons will most likely exceed 14. BAC (grey boxes), cosmid (white boxes) and long-range PCR product (black line) derived FISH probes enabled the positioning of the breakpoint (arrows indicate the relative direction of the breakpoint deduced from the presence/absence of the signals on the two derived chromosomes). Probes from BAC RPCI-11 89P5 and cosmids LA11197-C5, LA1163-H6, LA11236-G3 and LA1192-C6 indicated that the breakpoint was located near exons 2 and 3. A FISH probe synthesized from a long-range PCR product corresponding to the intronic sequence between these two exons indicated that the breakpoint lies upstream of the intron between exons 2 and 3.

Figure 15 5' sequence of the *GRIK4* gene showing the two possible N-terminal peptides derived from alternate start sites. Exon combination 1a-1a'-2 is derived from an EST sequence (acc. BE388730). Exon combination 1b-2 is based on the published cDNA sequence (e.g. acc. S67803). The actual amino acid sequence may differ from the published amino acid sequence as there is a potential downstream methionine start (MVAC... instead of MPRV...) containing a more conserved Kozak sequence (Kozak, 1986). It can be seen that the breakpoint upstream of exon 2 will separate the majority of the coding sequence from the promoter resulting in a putative null allele. Exonic DNA sequence is shown in capitals, intronic or upstream sequence in lower case. Conserved splice junction sequences (EXON/<u>GT</u>-----<u>AG</u>/EXON) are underlined. Single letter amino acid codes are shown beneath the appropriate DNA codons. A functional C/G:Leu/Val single nucleotide polymorphism (underlined) is found within exon 2.

Figure 16 shows the complete alternative nucleic acid sequence as identified by the present inventors.

Figure 17 shows the complete alternative amino acid sequence as identified by the present inventors.

Figure 18 shows the nucleic acid sequence of *NPAS3* spliceform 1.

Figure 19 shows the protein sequence of NPAS3 spliceform 1.

Figure 20 shows the nucleic acid sequence of *NPAS3* spliceform 2.

Figure 21 shows the protein sequence of NPAS3 spliceform 2.

Figure 22 shows an ideogram representation of the balanced translocation in patient 3 relating to this invention.

Figure 23 shows the genomic arrangement of the *NPAS3* gene including the position of the observed breakpoint.

Figure 24 shows potential functional consequences of the disruption to *NPAS3* gene : dominant-negative activity.

Figure 25 shows the *PDE4B1* nucleic acid sequence.

Figure 26 shows the PDE4B1 protein sequence.

Figure 27 shows the *PDE4B3* nucleic acid sequence.

Figure 28 shows the PDE4B3 protein sequence.

Figure 29 shows the *PDE4B2* nucleic acid sequence.

Figure 30 shows the PDE4B2 protein sequence.

Figure 31 a) Ideogram representation of balanced translocation between chromosomes 1 and 16 in patient 4.

Figure 32 Genomic arrangements of the *PDE4B* gene disrupted in the subject (patient 4). The two long transcripts of the *PDE4B* gene are shown. FISH showed the breakpoint was within a gap in the genome sequence between BACs RPCI-11 433N2 and RPCI-11 442I1. This positioned the breakpoint between the first and second exons of the *PDE4B1* form of the gene (acc. L20966). A long-range PCR product FISH probe corresponding to the genomic region encompassing the 1a exons of *PDE4B1* confirmed that the gene was disrupted between exon pairs 1a and exon 2 (i.e. only

PDE4B1 transcripts are directly disrupted by the chromosome abnormality).

Figure 33 shows an ideogram diagram of the chromosomal rearrangement (a reciprocal translocation) in patient 4. The two breakpoints are marked at the approximate chromosomal locations at which they are located. In addition, and not to scale, the two candidate disease-causing genes, *PDE4B* and *CDH8*, are placed in the correct orientation and with respect to the breakpoints. The fusion genes on derived chromosomes 1 and 16 that result from the reciprocal translocation are also indicated, demonstrating the potential capacity for fusion transcript/protein synthesis.

Figure 34 shows a representation of the genomic structure of the *CDH8* gene: its spliced exons spread over a genomic extent of approximately 400kb.
Above the gene, the coding contribution of each exon to the CDH8 protein is indicated by bars and finely dashed lines. The domain structure of CDH8 protein is shown at the top of the figure. 'N' and 'C' refer to the N- and C-termini of the protein. The broken line at the N-terminus indicates the existence of signal peptide and proprotein domains - both of which are cleaved off in the mature protein. The 'CD' ovals represent the positions of the five extracellular <u>c</u>adherin <u>d</u>omains. The black box signifies the position of the hydrophobic stretch of amino acids that act as the membrane-spanning domain. The BAC clones used to identify the breakpoint location are included at the bottom of the figure together with the inferred direction (arrows) of the breakpoint from the FISH results using these clones. The heavy dashed line shows the position of the breakpoint with respect to the gene exons and the domain structure of the protein.

Figure 35 Nucleic acid sequence of Human *CDH8*. The following features are marked for clarity:

a) ATG start site located at position 253 (underlined)

b) GC bases (underlined) at the junction between exons 1

and 2 (i.e. between which the breakpoint is located)

c) UGA stop codon located at position 2650 (underlined).

Figure 36 Amino acid sequence of Human CDH8 with underlined regions of interest.

a) Residues 1-29 signal peptide domain (italics)

b) Residues 30-61 propeptide fragment cleaved off in mature protein.

c) Residues 76-158 cadherin domain #1 (underlined)

d) Residues 172-248 cadherin domain #2 (underlined)

e) Residues 281-383 cadherin domain #3 (underlined)

f) Residues 396-487 cadherin domain #4 (underlined)

g) Residues 500-597 cadherin domain #5 (underlined)

h) 'V' highlighted at position 513 is the last residue in common with the putative truncated rat protein product from the alternatively spliced form.

i) Residues 622-645 transmembrane domain #1 (underlined).

Figure 37

a) Fusion protein product resulting from CDH8 promoter/exon 1 spliced to *PDE4B* exon 2 and beyond (transcribed on der(16)). The underlined residues 'RV' represent the fusion site between the two genes.

b) Fusion protein product resulting from *PDE4B* promoter (long form)/exon 1a spliced to *CDH8* exon 2 and beyond (transcribed on der(1)). See text for details: only the reading frame producing the N-terminal truncated form of the CDH8 protein is shown. The underlined 'gc' at position 68 represents the point of fusion between the two genes. Three potential methionine translation start sites are shown (highlighted) with the second of these having a nucleic acid sequence most similar to the canonical Kozak sequence (underlined). Use of this start site would generate a truncated CDH8 protein lacking the signal peptide, proprotein fragment, cadherin domain 1 and most of cadherin domain 2.

Materials and methods

Lymphocyte extraction and metaphase chromosome preparation

Lymphocytes were extracted from 7mls of patient blood (for storage and generation of EBV-transformed cell lines) using density gradient separation (Histopaque-1077, Sigma). In order to generate metaphase-arrested chromosomes for cytogenetic analysis, 0.8mls of patient blood were cultured for 71hrs in medium containing phytohaemagglutinin (Peripheral Blood Medium, Sigma). The short-term cultures were treated with colcemid for one hour followed by a conventional fixing procedure. Fixed chromosomes were dropped onto microscope slides and stored for 1 week prior to use in FISH experiments.

Selection of YAC clones for FISH probe synthesis

YAC clones were selected from the Whitehead/MIT map of the relevant chromosome in the cytogenetic intervals within which the breakpoints were adjudged to lie. YACs were obtained from the HGMP Resource Centre, Babraham Bioincubator, Babraham, Cambridge, UK (http://www.hgmp.mrc.ac.uk/). Clone DNA was prepared by standard methods and PCR amplified using primers designed against consensus sequence elements within the archetypal *Alu* repeat, Breen et al, 1992. This "Alu-PCR" gives a representative spread of non-repetitive sequence over the full length of the YAC and generates a better FISH probe than native YAC DNA. Alu-PCR was performed using the Expand Long Template PCR kit (Roche). Cycling conditions:94°C - 45s, 55°C - 30s, 68°C - 8min: 35 cycles. 68°C - 10min final extension.

Fluorescence in situ hybridisation (FISH) protocol

Probe template DNA (pooled Alu-PCR products, BAC clone DNA, cosmid clone DNA or long-range PCR products) were labelled by nick translation and hybridised to patient metaphase spreads using standard FISH methods. Slides were counterstained with DAPI in Vectashield anti-fade solution

(Vector laboratories). A Zeiss Axioskop fluorescence microscope with a chroma number 81000 or 830000 multi-spectral filter set was used to observe the chromosomal hybridisations. Images were captured using Vysis SmartCapture extension running within IP Lab spectrum or digital Scientific SmartCapture imaging software. FISH signals observed on derived chromosomes dictated the selection of further clones required to "walk" towards the breakpoint. Breakpoint-spanning FISH probes have signals on a normal chromosome and on both derived chromosomes.

Resolution of breakpoint position

BAC clones corresponding to positive YAC regions were arranged into contigs by consulting the Washington University FPC (http://www.genome.wustl.edu/gsc/human/Mapping/index.shtml), UCSC GoldenPath Draft Human Genome Browser (http://genome.ucsc.edu/goldenPath/hgTracks.html) and Ensembl (http://www.ensembl.org/) databases. BAC clones were supplied by BACPAC Resources, Oakland, California, USA (http://www.chori.org/bacpac/). Clone selection was biased to gene-containing BACs. Once a breakpoint-spanning BAC was identified, the position of the breakpoint in relation to candidate gene exons was determined by FISH probes generated from chromosome-specific library cosmids (HGMP Resource centre) or precisely positioned, repeat element-free long-range PCR products (Expand long range PCR kit, Roche; see below for primer sequences). Cycling conditions:94°C - 45s, 52°C - 30s, 68°C - 11min: 35 cycles. 68°C - 15min final extension. Cosmids were isolated by probing the appropriate chromosome-specific library filters (HGMP-RC) with isotopically labelled exon-specific PCR products.

Example 1: Molecular characterisation of chromosomal disruption and identification of disrupted gene from patient 1

FISH experiments on chromosome 3p13 had narrowed the location of the breakpoint to a region including the large gene *SEMCAP3* (approximately 250kb genomic extent). Two BAC clones were selected from the tiling diagram of BAC clones placed on the human genome map backbone (June 2002 release of the 'BAC End Pairs' track on the UCSC Genome Browser; http://genome.cse.ucsc.edu/index.html?org=Human). These were RPCI-11 606p16 and RPCI-11 94j25. By FISH, these BAC clones flanked the breakpoint (the former translocated to the derived chromosome 8 and the latter remained on the derived chromosome 3). The position of these two BAC clones indicated that the breakpoint lay within the large (200kb) intron between exons 3 and 4 of the *SEMCAP3* gene (see Fig.2). Thus, the inventors inferred from these results that the *SEMCAP3* gene was directly disrupted by the 3p13 translocation event and, as such, is a candidate gene for the psychiatric disorder exhibited by the patient.

Semcap3 (semaphorin cytoplasmic domain-associated protein) was originally identified in mouse as a gene encoding a protein that interacts with M-semF/Sema4c. Two forms, 3A and 3B, were submitted to the public nucleic acid sequence database (Wang & Strittmatter, 1999) but have yet to be published. It appears that 3b may be an artifactual sequence as it displays deletions in the sequence. Sema3a is identical in structure to the predicted human gene, KIAA1095 and the inventors refer to this sequence as human *SEMCAP3*. The yeast two-hybrid screen that isolated Sema3a/b also identified Sema1 and Sema2 as genes encoding proteins which interact with the cytoplasmic tail of the SEMA4C protein (Wang et al., 1999).

The purpose of these screening experiments was to elucidate cytoplasmic interactors with the transmembrane receptor, SEMA4C. This protein belongs to a large group of signalling proteins described as 'semaphorins'. In the

brain, these proteins are thought to play important roles in brain development through their action on axonal guidance and growth cone stability. Inagaki et al., (1995) showed that Sema4C is expressed in the developing mouse brain. One proposed explanation for the origin of psychiatric disorders (including the disorder exhibited by the patient described here) is the incorrect development of the brain, particularly the connections, projections and neural networks between brain subregions. With this in mind, semaphorins, and the proteins that interact with them (such as the SEMCAPs), become attractive candidate genes for the psychiatric disorders.

It is suspected that the PDZ domains (see Fig.2) of the SEMCAP3 protein will be involved in protein-protein interactions (such as SEMA4C interaction) as they are in other proteins. The RING-finger domain of SEMCAP3 identifies it as belonging to a class of proteins known as ubiquitin ligases. Ubiquitin ligases specifically target proteins for ubiquitination and subsequent destruction in the proteasome pathway. Thus, SEMCAP3 may act to regulate the activity of other proteins (for instance, components of the semaphorin pathway) by targeting them for destruction. The ZF-TRAF/SINA domain is most likely an extension of the RING-finger domain.

Figure 2 shows that the breakpoint would end *SEMCAP3* transcription after the third exon on the derived chromosome 3 (there would still be one normal chromosome 3 and *SEMCAP3* gene remaining in each nucleus). If transcription occurs on the derived chromosome 3 then the resulting translated protein product would be truncated; lacking part of the first PDZ domain and all subsequent amino acids in the C-terminal direction. It remains to be investigated if the psychiatric disorder in this patient results from *N33* perturbation on one allele, the disruption of *SEMCAP3* on one allele, the generation of an aberrantly functioning truncated SEMCAP3 from one allele or a combination of these.

Pulver *et al.* (1995) detailed schizophrenia linkage to chromosome 3p (albeit telomeric to *SEMCAP3*). However, two further studies have failed to replicate these findings in different populations (Maziade *et al.*, 2001 & Hovatta *et al.*, 1998).

Example 2: Further molecular characterisation of chromosomal disruption and identification of disrupted gene

In this case, primers corresponding to N33 3'UTR sequences and an STS, SHGC-12093 (Acc. No. G17275) were designed (see below for primer sequences). These PCR products were used to screen the chromosome 8 specific cosmid library (LA08). Among others, positive cosmids LA0854-H5 (3' UTR) and LA08145-E3 (STS) were isolated and subsequently used in FISH experiments (see below for results).

3'UTR primers
Primer A: TGCCACGTGTTAGCAGAAAG
Primer B: TGCCTTTAACCAGATGAGGC

SHGC-12093 primers
Primer A: TCTTGTGGGTCACAATTAGGC
Primer B: TAAAAAGGTGCAGTTTCTTCAGC'.

The subject has schizoaffective disorder and a balanced reciprocal translocation between chromosomes 3 and 8. A 8p22 breakpoint-crossing YAC, 931_a_1, was identified. This permitted a 8p22 breakpoint-crossing BAC RPCI-11 23j14 (acc. no. AC019292) to be found. This was shown to contain the 3' end of the *N33* gene (Fig.6). Subsequently, FISH with cosmids LA0854-H5 and LA08145-E3 from the LANL chromosome 8 specific library (HGMP Resource Centre, Babraham, Cambridge, UK) flanked the breakpoint, placing it approximately 100Kb from exon 11 of N33. N33 is related to a number of genes, human IAG2, Drosophila CG7830, C.

elegans g304348 and two yeast proteins, OST3 and OST6 (see Fig. 8 for alignment of proteins). While the homologies between N33 and the yeast proteins are relatively weak, they share conserved cysteine residues and have the same locations for the four transmembrane domains as predicted by hydropathy plots. Ost3 and Ost6 are components of the oligosaccharyl transferase complex responsible for the addition of oligosaccharides to selected proteins. This has been backed up by protein structure prediction programs detailed in a recent report Fetrow et al, 2001.

The present inventors have identified an alternative start exon, herein identified as exon 1a (see Figures 5 & 6) to that in the public database, herein identified at exon 1b. Additionally they have identified a complex variation of splicing with the exons and proposed sequences of the transcripts, shown in Figures 5, 6 and 37 respectively. In view of the complex splice variations the C-terminal sequence of the various N3 splice forms is predicted to vary and this is shown in Figure 9.

Because N33 lies within a linkage hotspot for schizophrenia (Gurling et al, 2001, Brzustowicz et al, 1999, Blouin et al, 1998, Kaufmann et al, 1998, Kendler et al, 1996, Pulver et al, 1995) the present inventors decided to carry out an association study on this gene. Three microsatellite markers (D8S549, N33 microsatellite and D8S1992

Microsatellites used in associated study

D8S549

Primer A: AAATGAATCTCTGATTAGCCAAC

Primer B: TGAGAGCCAACCTATTTCTACC

N33 microsatellite

Primer A: AGGCTGAGTGCCAAAAAGTA

Primer B: CTTTAAGCTTGCTATTTGAAGGC

D8S1992

Primer A: TTCATCGTCTGAACCTGG

Primer B: ACACATTTCCTCTATGTTGC) were chosen and used to type 25 mother-father-schizophrenic proband trios and 64 schizophrenic cases and 64 normal controls. The haplotypes derived from the trio study were examined for frequency bias in the case and control samples. Certain haplotypes are currently over-represented in the schizophrenic case genotypes compared to controls. Appropriate individuals with the haplotypes are currently being screened for mutations.

Example 3: Molecular characterisation of chromosomal disruption and identification of disrupted gene from patient 2

Psychiatric evaluation

The subject (female) was approached and gave full, informed written consent for this study as one of a large cohort of people co-morbid for schizophrenia and mental retardation. Prior to investigation she was not known to have any abnormality of karyotype. She suffered from chronic schizophrenia and a mild degree of mental retardation (IQ between 65-70). The diagnosis of chronic schizophrenia was confirmed using SADS-L structured interview to generate DSM-IV and ICD-10 criteria, by a psychiatrist experienced in both general psychiatry and the psychiatry of mental retardation (WM). SADS can be reliably used in patients with mild mental retardation. Consensus diagnosis was reached on review by two psychiatrists (WM and DB). IQ scores were generated from WAIS-R and their stability shown by similar levels detected by psychological examination at different times throughout her life. There were no dysmorphic features in the subject. However the subject did suffer from bilateral deafness since childhood - a consequence of surgical operations on the mastoids. There was no family history of mental illness or mental retardation that could be

ascertained. Other members of the family declined to participate in the study.

An initial G-banded karyotype of this patient indicated that the chromosome abnormality was complex (46, XX,ins(8;11)(q13;q23.3q24.2)inv(2)(p12q32.1) t(2;11)(q21.3;q24.2)der(2)(2qter->2q32.1::2p12->2q21.3::11q24.2->11qter) der(11)(11pter->11q23.3::2q21.3->2q32.1::2p12->2pter)der(8)(8pter->8q13::11q23.3->11q24.2::8q13->8qter)), involving a pericentric inversion of chromosome 2 coupled with rearrangements involving chromosomes 2, 8 and 11 (Fig.13). Figure 12 details the YAC and BAC FISH probes crossing or bracketing breakpoints on 2 and 11. Sequence in the locality of the breakpoints was assessed for gene content.

PCR primers

Long-range PCR for FISH probe templates:
Int2-3 GRIK4a; CAGGAGGTCCTGTGAAGCTC,
Int2-3 GRIK4b; ACAGGGAAAGAAGCAAAGCA.

GRIK4 exon region-specific PCR: screening of chromosome 11 cosmid libraries:

Ex1a/a' a; AAAGCTAAGCGCAGGTGTGT,
Ex1a/a' b; TTTCTGGGAGGCAACCATAG,
Ex1b a; GCAGAGTTATGTCATGCCCA,
Ex1b b; CCTGTGCAGCACTCTGATGT,
Ex2/3 a; TTGAACCCAAGAGAACAGGG,
Ex2/3 b; TCCCCTTCTCCTTCCAGTTT
Cycling conditions:94°C - 2min initial denaturation. 94°C - 1min, 52°C - 1min, 72°C - 75s: 33 cycles. 72°C - 15min final extension.

The 11q23.3 breakpoint is located at a locus containing a kainate-type ionotropic glutamate receptor (*GRIK4,* acc. S67803 & NM_014619 (11), previous nomenclature *KA1/EAA1*). Cosmid FISH directed at the individual exons and an intron-specific long-range PCR product FISH (Fig.15)

positioned the breakpoint within the *GRIK4* gene sequence; most likely immediately upstream of exon 2 (our nomenclature, Fig.15). This was confirmed using a long-range PCR product FISH probe corresponding to the intron between exons 2 and 3 (Fig.15). We also identified a GenBank EST (acc. BE388730, IMAGE clone ID:3613199) generating an alternative start-site resulting in an alternative cognate N-terminal peptide sequence (Figures 16 and 17). The position of a breakpoint anywhere between exons1a/a'/1b and exon 3 would truncate all putative transcript forms such that no receptor function could be encoded on the derived chromosome 11. Hence, the patient had only one intact *GRIK4* allele.

Discussion

The present inventors identified a subject with comorbin schizophrenia with mild learning disability in whom chromosome translocation events have disrupted brain-expressed gene that are also functional disease candidates. Without wishing to be bound by theory it is hypothesised that the disruption of the *GRIK4* gene by a chromosomal breakpoint (and the resulting reduced gene dosage) is the principal underlying cause of psychiatric disease in this patient.

The gene disrupted in this patient is both expressed in the brain and participates in key physiological processes in the CNS. Notably, the gene may be involved in the alteration of the strength of synaptic/neural transmission, a phenomenon known as long-term potentiation (LTP). LTP is postulated to underlie cognitive functions such as learning and memory. Moreover, cognitive testing has previously established that these functions are frequently affected in patients with schizophrenia.

GRIK4

Three classes of ionotropic glutamate receptors have been identified on the basis of their pharmacological profiles and sequence homologies; NMDA receptors, AMPA receptors and Kainate receptors. Functional Kainate receptors *in vivo* may be heteromeric, consisting of combinations of the low kainate agonist affinity (GLUR5, GLUR6 and GLUR7) and high-affinity subunits (GRIK4 and GRIK5) (Chittajallu et al, 1999; Lerma et al, 2001 and Werner et al, 1991). The subject with comorbid schizophrenia and mild learning disability possesses a complex chromosomal rearrangement. Of all the breakpoints studied in this patient only the *GRIK4* gene is directly disrupted. This might be expected to modify kainate receptor channel properties by altering subunit stoichiometry.

The glutamate receptors are key initiators of synaptic LTP (Miller and Mayford, 1999). NMDA receptors are the principal mediators of LTP but recently presynaptic kainate receptor-dependent plasticity changes have been observed at mossy fibre synapses in the hippocampus (Contractor et al, 2001 and Lauri et al, 2001). Interestingly, an involvement of the glutamate neurotransmitter system in the pathophysiology of schizophrenia has been postulated. The "Glutamate Hypothesis" attempts to explain the psychotic symptoms that arise following administration of ionotropic glutamate receptor antagonists such as phencyclidine (PCP; "Angel Dust") and ketamine (Goff and Nine, 1997). Several studies also point to changes, predominantly decreases, in glutamate receptor subunit expression (including kainate receptors) in the brains of schizophrenic patients (Ibrahim et al, and Meador-Woodruff, 2001). Similarly, Mohn et al, 1999 report that mutant mice with reduced NMDAR1 (another glutamate receptor) expression levels display schizophrenia-like behaviours.

As well as aberrant neurotransmission function in the adult, it has been suggested that neurodevelopmental deficits may contribute to schizophrenia. Neuroanatomical studies indicate statistically significant reduced volumes of brain regions, primarily the hippocampus, in schizophrenic and comorbid patients (Sanderson et al 1999 and Pearlson, 1999). *GRIK4* is expressed in the amygdala, hippocampal formation (CA3 pyramidal and dentate granule cells) and entorhinal cortex. Glutamate receptors might mediate brain development through the activity-dependent refinement of neuronal connections.

The present subject was clinically diagnosed as having schizophrenia coupled with mild learning disability. It may be the case that causative gene mutations in comorbid patients lead to a severe phenotype or have more profound downstream effects than gene mutations in patients with schizophrenia alone (i.e. the comorbid state represents the severest form of schizophrenia (Doody et al, 1998)). A second possibility is that the gene mutation gives rise to the learning disability component of the illness through an independent effect on brain development. The manner in which the mutated genotype gives rise to the observed phenotype (via functional or developmental mechanisms) is a key issue in molecular neurobiology, particularly in the characterisation of mouse "knockout" mutants (Mayford et al, 1995).

A large number of publications detail family and population-based linkage studies carried out to identify psychiatric illness susceptibility loci. The results have not been conclusive perhaps indicating the presence of confounding factors such as population stratification, incomplete penetrance, genetic heterogeneity and uncertain mode of inheritance. Nevertheless, *GRIK4* lies at the edge of a schizophrenia linkage region described in a recent publication (Gurling et al, 2001). The most centromeric marker exhibiting linkage to schizophrenia in this paper,

D11S925, is located within an intron at the 3' end of *GRIK4*.

Example 4: Molecular characterisation of chromosomal disruption and identification of disrupted gene from patient 3

Fine FISH mapping of the breakpoint with cosmid clones

PCR products corresponding to regions in or near h*NPAS3* exons 4, 5 and 6 were obtained using the following primers under standard PCR conditions (Exon 4-*i* ACAACCATTCTGGGAACAGC, Exon 4-*ii* GTGTAGGGAAAGCCATCCAA, Exon 5-*i* TCTTTTTCCTGCAGTCCCTG, Exon 5-*ii* CTCCAAATGACTCCTGCCAT, Exon 6-*i* GCCTCTGCCATAGATTTTGC, Exon 6-*ii* TTCCTTCCCACCCTTTCTCT). Probes were created by random-primed labelling of PCR products with radioactive dCTP; these were used to screen a LANL chromosome 14-specific cosmid library (LA14NC01 obtained from the UK HGMP Resource Centre, Hinxton, Cambridge) using hybriding conditions set out in Church and Gilbert (1986). Positive clones (exon; LA1431-G5, exon 5: LA14123 - C4 and exon 6; LA1487 - D9) were prepared by a standard alkaline lysis protocol and taken through FISH analysis as above.

__Results__

Metaphase spreads from EBV-transformed cell lines were analysed by Fluorescence in situ Hybridisation (FISH) using successively smaller DNA probes. A breakpoint spanning BAC clone was obtained by FISH screening (RPCI-11 BAC 1078i14, acc. no. AL161851). EST sequences were examined in the genomic DNA flanking the breakpoint in order to identify potential transcripts in the locality. A number of ESTs were identified which had been annotated as containing homologous sequence to the conserved "PAS" domain present in a large number of genes (Gu et al, 2000). A search of such genes revealed that the most closely related gene encoded a mouse brain-expressed transcript, neuronal pas domain protein 3 (NPAS3 (MOP6), acc. no. AF137871;

hereafter referred to as mNPAS3). Nucleotide homology to the mNPAS3 cDNA within human genomic DNA BAC clone sequences at 14q13 using the BLAST algorithm identified 12 exons corresponding to the human orthologue of mNPAS3 (*hNPAS3*) distributed over a genomic region of approximately 800-900Kb making it among the largest gene loci in the human genome (Figure 23). Subsequently, full length h*NPAS3* cDNA sequences have been submitted by two other groups to GenBank/EMBL with the accession numbers, AB054575 and AF164438, although these have differences to the mouse splice-form in the 5' exons. This is due to the presence of two alternative transcription start sites employed in both human and mouse genes. This was confirmed by analysis of published cDNA and EST sequences coupled with further sequencing of corresponding IMAGE clones. These splice variants are highlighted in Figures 18, 30 and 23.

The ratio of fluorescent signals on the derived chromosomes 9 and 14 from the breakpoint-spanning BAC probe, 1078i14, indicated that the breakpoint was located at the centromeric end of the BAC. This is the location of exon 5 of the gene. Exon 4-, 5- and 6-containing cosmids were isolated and used as FISH probes to provide definitive proof of the location of the breakpoint and confirmation that a full-length transcript (and hence protein) cannot be synthesized on the derived chromosome 14. An exon 5-containing cosmid (see Figure 23) spanned the breakpoint. Subsequently a long-range PCR product-derived FISH probe corresponding to exon 5 indicated that the breakpoint lay upstream of exon 5.

Long-range PCR primers - NPAS3 exon 5

a) ccagcttgtatgtggtgtgg

b) ttactcccagtgcccattgt.

<u>Discussion</u>

A FISH-based approach has shown that the gene, *NPAS3*, is disrupted by a chromosomal rearrangement present in a mother and daughter who suffer from comorbid schizophrenia and learning disability respectively. NPAS3 is a brain expressed transcription factor of the basic helix-loop-helix PAS domain class which includes members such as AHR and ARNT.

Neuronal pas3 (NPAS3) was originally cloned in the mouse (Brunskill et al, 1999) on the basis of its sequence homology with other PAS domain proteins. Its expression has been characterised in the developing mouse embryo where high levels are seen in the neural tube, neuroepithelium and, later, the neopallial layer of the cortex. Non-neural expression was also observed in the heart, limb and kidney. In the mouse, NPAS1 (human chromosomal location, 19q13) is expressed in deep pyramidal cortex cells, hippocampus and amygdala (Zhou et al., 1997). NPAS2 (human chromosomal location, 2q13) is expressed in the cortex, hippocampus and thalamus. Lower levels were also seen in spinal cord, intestines and uterus. NPAS2 was also recently deleted in mice by homologous recombination (Garcia et al., 2000) leading to deficits in cued and contextual memory. In addition NPAS2 appears to have a role in cellular energy state monitoring and the circadian rhythm pathway (Reick et al, 2001 and Rutter et al, 2001). The translocation event described herein disrupts the gene between exons 4 and 5. If transcription occurred at this disrupted locus, a truncated protein would result containing only the bHLH domain. It is conceivable that this protein would have a dominant negative effect on wild-type NPAS3 protein (or any other heterodimeric protein partner) through the creation of non-functional dimers (see Figure 24 for explanatory diagram). This would result in a potentially more severe or penetrant phenotype than a conventional point mutation. Two examples where bHLH-PAS proteins have been altered through loss of the C-terminal PAS domain (one

experimentally, the other in a patient with a chromosome translocation) have resulted in probable dominant negative action (Maemura et al, 1999, Holder jr. et al, 2000).

Mutations in this gene in karyotypically normal individuals would not be expected to have as severe or penetrant effects as those observed in the two t(9;14) patients.

Sequence comparison between hNPAS3 and other members of the NPAS sub-family show that homologies are largely restricted to the N-terminal end of the protein; the location of a basic helix-loop-helix and PAS domains. The greatest homology is with NPAS1, then NPAS2 and other PAS domain-containing proteins (data not shown). An alignment of the cognate human (conceptually translated from the splice-form containing exons 1-12) and mouse NPAS3 proteins reveals near identity over the N-terminal half of the protein but increased divergence at the C-terminal end. This is particularly the case for two stretches where 5 and 7 amino acids, respectively, have been gained in the human orthologue (Fig.21). These correspond to two poly-glycine tracts present within exon 12 (of 11 and 10 residues respectively). Such tracts can be indicative of slipped strand mispairing whereby trinucleotide repeats are aberrantly expanded or deleted. Where they occur in coding sequence, increases in the number of trinucleotide repeats can have a pathological effect on protein function (e.g. Huntington disease and Spino-cerebellar ataxia 1). Another feature of such repeats is their unstable nature between generations: a lowering of the age of onset of a disease from generation to generation (anticipation) can often be directly linked to an increase in the number of repeat units.

Exon 12 (coding for the C-terminus of the protein) is also noteworthy because of the extremely high density of CpG dinucleotides (in humans and mouse); a feature that abruptly ends at the junctions with flanking intronic/3' sequences. This "CpG island" is unusual because it is both

transcribed and also located at the 3' rather than 5' end of the gene. The significance of this in terms of potential transcriptional control by methylation or susceptibility to mutation is as yet unknown. However, the high level of G and C bases creates a bias in amino acid composition such that alanine, glycine, histidine and proline are over-represented. This may explain the presence and expansion of the poly-glycine tracts in Npas3.

14q13 is also the site of linkage to Fahr's syndrome (idiopathic basal ganglia calcification; IBGC) as determined from analysis of families (Geschwind et al, 1999). Fahr's syndrome symptoms are often accompanied by psychoses such as schizophrenia. Thus, it may be the case that NPAS3 is also the gene responsible for Fahr's syndrome.

Example 5: Molecular characterisation of chromosomal disruption and identification of disrupted gene from patient 4

Psychiatric evaluation

The subject (male) is the proband in a family segregating a t(1;16) balanced reciprocal translocation. He gave full informed consent to the study. His diagnosis of chronic schizophrenia was confirmed by SADS-L structured interview and a consensus reached by two psychiatrists (WM and DB). He does not have mental retardation. Other members of his near family also gave consent to participate in this study, none of whom had current mental illness (several are below the age of risk for psychiatric illness). There was also a history of mental illness (major depressive disorder) in members of the extended family who were known to be translocation carriers, but they could not be approached for confirmation at the time of the current study. An unrelated individual (now deceased) with DSM-IV chronic schizophrenia without learning disability also had a t(1;16) balanced

translocation with the same breakpoints (at the resolution of G-banding).

PCR primers

Long-range PCR for FISH probe templates: PDE4B3a; GTCAGACAAATCCAAATGGAGAG, PDE4B3b; CTTTCTCCTGTCACTTTCCTTCA.

Cycling conditions:94°C - 2min initial denaturation. 94°C - 1min, 52°C - 1min, 72°C - 75s: 33 cycles. 72°C - 15min final extension.

The balanced translocation, t(1;16)(p31.2;q21), in this family results in two breakpoints (Figure 33). Genomic sequence at 16q21 is not complete. The only known gene in the vicinity of the breakpoint region is Cadherin 8 (*CDH8*, acc. AB035305).

In contrast, on chromosome 1p31.2 FISH identified two non-overlapping BAC clones (RP11-433N2, acc. AL513493 and RP11-442I1, acc. AL391359) which reside on either side of the breakpoint in this patient. The breakpoint-containing genomic region between these two BAC clones has yet to be sequenced (see Figure 32). Database annotation of the two BAC clones together with BLAST mapping of exons onto genomic sequence indicated that this locus contains a cAMP phosphodiesterase gene, *PDE4B*. Two cDNAs corresponding to longer transcript forms of this gene (denoted *PDE4B1*, acc. L20966 and *PDE4B3*, acc. U85048, respectively) have been previously characterised (Bolger et al, 1994; Huston et al, 1997). Long-range PCR product FISH (Figure 32) confirmed that the *PDE4B1* transcript is directly disrupted by the breakpoint (although additional position-effect perturbation of *PDE4B3* expression cannot be ruled out). Huston et al. (1997) have previously shown that the *PDE4B1* transcript encodes an alternative N-terminal peptide sequence. In addition, they demonstrated that only this form is expressed in the brain. It is therefore predicted that this patient will have a reduction in the levels of functional PDE4B in the brain.

Discussion

The present inventors have identified a subject with DSMIV chronic schizophrenia in whom chromosome translocation events have disrupted brain-expressed genes that are also functional disease candidates. Without wishing to be bound by theory it is hypothesised that the disruption of the *PDE4B* gene by a chromosomal breakpoint (and the resulting reduced gene dosage) is the principal underlying cause of psychiatric disease in this patient.

The gene disrupted in this patient is both expressed in the brain and participates in key physiological processes in the CNS. Notably, the gene may be involved in the alteration of the strength of synaptic/neural transmission, a phenomenon known as long-term potentiation (LTP). LTP is postulated to underlie cognitive functions such as learning and memory. Moreover, cognitive testing has previously established that these functions are frequently affected in patients with schizophrenia.

PDE4B

Stimulation of the G protein coupled receptor/heterotrimeric G protein pathway results in the synthesis of the secondary messenger, cAMP, by members of the adenylyl cyclase family of enzymes. This secondary messenger triggers a well-characterised signalling cascade that is principally mediated by cAMP-dependent protein kinase A (PKA) and cAMP-resposive transcription factor, CREB, both of which have been implicated in the molecular pathways of LTP (Abel & Latal, 2001). cAMP signalling is attenuated by its breakdown by members of the phosphodiesterase enzyme family. Four members of the *PDE4* sub-family of cAMP phosphodiesterases have been identified to date (*PDE4A-PDE4D*). These four genes are the human homologues of the *Drosophila* learning and memory mutant gene, *Dunce*. The long form of the PDE4B protein, PDE4B1, is the only splice form with brain expression and the present inventors have shown that it is disrupted in the

subject. Anti-PDE4B antibodies revealed expression within the inferior olive, the hypothalamus, the ventral striatum, the cerebellar molecular layer, globus pallidus, nucleus accumbens and substantia nigra (Cherry & Davis, 1999). The authors of this expression study suggested that *PDE4B* expression strongly correlates with brain areas underlying reward and affect in mammals. In addition, PDE4 proteins are recognised as the molecular targets for Rolipram, a drug with anti-depressant effects. Rolipram inhibition of PDE4 activity has been shown to improve long-term hippocampal LTP and spatial memory in mice (Barad et al, 1998 and Bach et al, 1991). The (heterozygous) disruption to *PDE4B1* described here may be equivalent to 50% reduction of protein product in the brain. This could result in a greater cAMP half-life and a concomitant increase in the activation of downstream cAMP targets.

In addition, the disruption to *PDE4B* shows reduced penetrance as not all translocation carriers present with psychiatric illness (although all members of the extended family with psychiatric illness possess the translocation karyotype; data not shown).

Example 6: Molecular characterisation of chromosomal disruption and identification of disrupted gene from patient 4

FISH experiments on chromosome 16q21 had narrowed the location of the breakpoint to a region including the large gene *CDH8* (approximately 400kb genomic extent). Three BAC clones were selected from the tiling diagram of BAC clones placed on the human genome map backbone (June 2002 release of the 'BAC End Pairs' track on the UCSC Genome Browser; http://genome.cse.ucsc.edu/index.html?org=Human). These were RPCI-11 599c11, RPCI-11 875e12 and RPCI-11 685m21. By FISH, these BAC clones flanked the breakpoint (the first two translocated the derived chromosome 1 whereas the third remained on the derived chromosome 16). The position of these three BAC clones indicated that the breakpoint lay

within the large (100kb) intron between exons 1 and 2 of the *CDH8* gene (see Fig.2). Thus, the inventors inferred from these results that the *CDH8* gene was directly disrupted by the 16q21 translocation event and, as such, is a candidate gene for the psychiatric disorder exhibited by the patient. The similar disruption of the *PDE4B* gene on chromosome 1 and their relative orientations on the two chromosomes raised the possibility that the derived chromosomes (the two chromosomes resulting from the translocation: der(1) and der(16)) could transcribe fusion/hybrid genes. This has been frequently seen in cases where a translocation gives rise to susceptibilty to cancers. In essence, the translocation in the proband resulted in an exchange of the two genes' promoter and first exon sequences. On the der(1) the promoter and first exon of the *CDH8* gene are juxtaposed to exon 2 and downstream of the *PDE4B* gene (see Fig.33). However, the reading frames of these two gene segments are not the same, resulting in a prematurely truncated peptide with only the signal peptide, proprotein fragment and a small portion of the cadherin domain contained within (see Fig. 37a). This would be expressed in the same cell types/tissues as the normal *CDH8* gene but the functional/pathological significance of this small peptide is not clear at the current time. On the der(16) the *PDE4B* promoter and exon 1a are juxtaposed to exon 2 and downstream of the *CDH8* gene (see Fig.33). Exon 1a of *PDE4B* does not contain a translation start-site so the reading frame compatabilities of the putative fused transcript are not an issue. However, exon 2 and downstream of the *CDH8* gene contain several ATG start-sites which could be employed by translational machinery to generate peptide sequences. In two of the reading frames, any generated peptides would be small and probably of no consequence. The third reading frame (the normal *CDH8* reading frame, see Fig.5b) contains three ATG start-sites early on, with the second of these forming a very good match to the canonical Kozak sequence found at

most translation start-sites (CCAxx<u>ATG</u>G). If this one is used then the resulting peptide will be identical to normal CDH8 protein but lacking the N-terminal portion encoding the signal peptide, proprotein fragment, the first cadherin domain and most of the second cadherin domain. Although the bulk of the peptide sequence is as the normal CDH8 protein, the lack of the N-terminal sequences may prevent the protein from entering the Golgi/ER subcellular compartments - a process that is required for the correct insertion in/trafficking to the cell membrane. The functional/pathological consequence of the presence of this truncated form of the CDH8 protein in the cytoplasm of tissues where the long form of the *PDE4B* gene is expressed is uncertain at this point.

In summary, the psychiatric illness seen in the proband, and other members of the family, may be the result of one (or a combination) of the following circumstances: the loss (through disruption) of one allele of *PDE4B*, the loss (through disruption) of one allele of *CDH8* or the generation of potentially pathological fusion polypeptides.

Cadherin-8 was first cloned in humans (Tanihara et al., 1994) and later in mouse (Munro et a., 1996) and rat (Kido et al., 1998). Sequence analysis immediately placed the gene product within the large family of membrane-spanning proteins with extracellular cadherin domains thought to mediate calcium-dependent homophilic interactions between adjacent cells. As such, the cadherins are members of the functionally defined group of cell adhesion proteins.

CDH8 is a member of the Type II, or atypical, cadherins which are defined by the lack of an extracellular tripeptide motif, HAV, possibly involved in the binding specificity of Type I cadherins. Fig.2 illustrates the structure of CDH8 protein which includes an extracellular domain containing 5 copies of the cadherin domain, a membrane spanning domain and a C-terminal cytoplasmic tail. The cytoplasmic tail is thought to signal the presence of

interactions to the intracellular compartment by mediating receptor clustering through interaction with the proteins such as β-catenin, α-catenin and, eventually, the cytoskeletal proteins, actin and α-actinin. In this way, adhesion to adjacent cells can affect the cytoarchitecture of the cell and may even play a role in cell motility.

The two principal roles of neuronal cadherins are thought to be in the mediation of certain developmental pathways in the brain and the regulation of synaptic function. The homophilic nature of cadherin interaction (i.e. CDH8 proteins preferentially bind to other CDH8 proteins) has prompted the hypothesis that cadherins are responsible for the aggregation or interconnection of similar cells within an organ. This has been shown to be the case in the brain where *CDH8* expression has been shown to be restricted to particular subregions and even neuronal patches (Redies, Bishop, Rubenstein, Korematsu X 2).

The major cadherin in the brain, N-cadherin (encoded by *CDH2*), has been implicated in synaptic long-term potentiation (LTP): the mechanism thought to underlie learning and memory on the brain (e.g. Huntley et al., 2002 & Bozdagi et al., 2000,). Other cadherins may also play a part in this process (Uemura, 1998 & Tang et al., 1998). In essence, cadherins seem to form physical bridges across the synaptic cleft which may modify synaptic efficacy and/or spine morphology (two features of neurons demonstrated to change after the induction of LTP).

Interestingly, two of the hypotheses used to explain the origins of psychiatric illness are, firstly, the occurrence of abnormal brain development and, secondly, the existence of deficits in cellular pathways manifested as poor performance in certain cognitive/memory tasks. The two roles of neuronal cadherins seem to closely mirror these two hypotheses suggesting that CDH8 is a good functional candidate for psychiatric illness.

REFERENCES

1. Berrettini W.H. (2000) Genetics of psychiatric disease. Annu. Rev. Med. 51:465 - 79.

2. Lai C.S, Fisher S.E, Hurst J.A, Vargha-Khadem F, Monaco A.P. (2001) A forkhead-domain gene is mutated in a severe speech and language disorder. Nature. 413(6855):519 - 23.

3. Kleinjan D.J, van Heyningen V. (1998) Position effect in human genetic disease. Hum Mol Genet. 7(10):1611 - 8.

4. Brennan J, Skarnes W.C. (1999) Gene trapping in mouse embryonic stem cells. Methods Mol. Biol. 97:123 - 38.

5. Breen M, Arveiler B, Murray I, Gosden J.R, Porteous D.J. (1992) YAC mapping by FISH using Alu-PCR-generated probes. Genomics. 13(3):726 - 30.

6. Fetrow J.S, Siew N, Di Gennaro J.A, Martinez-Yamout M, Dyson H.J, Skolnick J. (2001) Genomic-scale comparison of sequence- and structure-based methods of function prediction: does structure provide additional insight? Protein Sci. 2001 May;10(5):1005 - 14.

7. Gurling H.M, Kalsi G, Brynjolfson J, Sigmundsson T, Sherrington R, Mankoo B.S, Read T, Murphy P, Blaveri E, McQuillin A, Petursson H, Curtis D. (2001) Genomewide genetic linkage analysis confirms the presence of susceptibility loci for schizophrenia, on chromosomes 1q32.2, 5q33.2, and 8p21-22 and provides support for linkage to schizophrenia, on chromosomes 11q23.3-24 and 20q12.1-11.23. Am J Hum Genet. Mar;68(3):661 - 73.

8. Brzustowicz L.M, Honer W.G, Chow E.W, Little D, Hogan J, Hodgkinson K, Bassett AS. (1999) Linkage of familial schizophrenia to chromosome 13q32. Am J Hum Genet. Oct;65(4):1096 - 103.

9. Blouin J.L, Dombroski B.A, Nath S.K, Lasseter V.K, Wolyniec P.S, Nestadt G, Thornquist M, Ullrich G, McGrath J, Kasch L, Lamacz M, Thomas M.G, Gehrig C, Radhakrishna U, Snyder S.E, Balk K.G, Neufeld K, Swartz K.L, DeMarchi N, Papadimitriou G.N, Dikeos D.G, Stefanis C.N, Chakravarti A,

Childs B, Pulver A.E, et al. (1998) Schizophrenia susceptibility loci on chromosomes 13q32 and 8p21. Nat Genet. Sep;20(1):70 - 3.

10. Kaufmann C.A, Suarez B, Malaspina D, Pepple J, Svrakic D, Markel P.D, Meyer J, Zambuto C.T, Schmitt K, Matise T.C, Harkavy Friedman J.M, Hampe C, Lee H, Shore D, Wynne D, Faraone S.V, Tsuang M.T, Cloninger C.R. (1998) NIMH Genetics Initiative Millenium Schizophrenia Consortium: linkage analysis of African-American pedigrees. Am J Med Genet. Jul 10;81(4):282 - 9.

11. Kendler K.S, MacLean C.J, O'Neill F.A, Burke J, Murphy B, Duke F, Shinkwin R, Easter S.M, Webb B.T, Zhang J, Walsh D, Straub (1996) Evidence for a schizophrenia vulnerability locus on chromosome 8p in the Irish Study of High-Density Schizophrenia Families. Am J Psychiatry. Dec;153(12):1534 - 40.

12. Pulver A.E, Lasseter V.K, Kasch L, Wolyniec P, Nestadt G, Blouin J.L, Kimberland M, Babb R, Vourlis S, Chen H, et al. (1995) Schizophrenia: a genome scan targets chromosomes 3p and 8p as potential sites of susceptibility genes. Am J Med Genet. Jun 19;60(3):252 - 60.

13. MacGrogan D, Levy A, Bova G.S, Isaacs W.B and Bookstein R. Structure and methylation-associated silencing of a gene with a homozygously deleted region of himan chromosome band 8p22. Genomics 35 (1), 55 - 65 (1996).

14. Brennan J, Skarnes W.C. (1999) Gene trapping in mouse embryonic stem cells. Methods Mol. Biol. 97:123 - 38.

15. Doody G.A, Johnstone E.C, Sanderson T.L, Owens D.G, Muir W.J. (1998) "Pfropfschizophrenie" revisited. Schizophrenia in people with mild learning disability. Br J Psychiatry. 173:145 - 53.

16. Sanderson T.L, Doody G.A, Best J, Owens D.G, Johnstone E.C. (2001) Correlations between clinical and historical variables, and cerebral structural variables in people with mild intellectual disability and schizophrenia. J Intellect Disabil Res. 45(Pt 2):89 - 98.

17. Kamboj R.K, Schoepp D.D, Nutt S, Shekter L, Korczak B, True R.A, Rampersad V, Zimmerman D.M, Wosnick M.A. (1994) Molecular cloning, expression, and pharmacological characterization of humEAA1, a human kainate receptor subunit. J Neurochem. 62(1):1 - 9.

18. Martin S.J, Grimwood P.D, Morris R.G. (2000) Synaptic plasticity and memory: an evaluation of the hypothesis. Annu Rev Neurosci. 23:649 - 711.

19. Kuperberg G, Heckers S. (2000) Schizophrenia and cognitive function. Curr Opin Neurobiol.. 10(2):205 - 10.

20. Elvevag B, Goldberg T.E. (2000) Cognitive impairment in schizophrenia is the core of the disorder. Crit Rev Neurobiol. 14(1):1 - 21.

21. Blackwood D.H, Glabus M.F, Dunan J, O'Carroll R.E, Muir W.J, Ebmeier K.P. (1999) Altered cerebral perfusion measured by SPECT in relatives of patients with schizophrenia. Correlations with memory and P300.Br J Psychiatry. 175:357 - 366.

22. Chittajallu R, Braithwaite S.P, Clarke V.R, Henley J.M (1999) Kainate receptors: subunits, synaptic localization and function. Trends Pharmacol Sci. 20(1):26 - 35.

23. Lerma J, Paternain A.V, Rodriguez-Moreno A, Lopez-Garcia J.C. (2001) Molecular physiology of kainate receptors. Physiol Rev. 81(3):971 - 98.

24. Werner P, Voigt M, Keinänen K, Wisden W, Seeburg P.H (1991) Cloning of a putative high-affinity kainate receptor expressed predominantly in hippocampal CA3 cells. Nature. 351:742 - 4.

25. Miller S, Mayford M. (1999) Cellular and molecular mechanisms of memory: the LTP connection. Curr Opin Genet Dev. 9(3):333 - 7.

26. Contractor A, Swanson G, Heinemann S.F. (2001) Kainate receptors are involved in short- and long-term plasticity at mossy fiber synapses in the hippocampus. Neuron. 29(1):209 - 16.

27. Lauri S.E, Bortolotto Z.A, Bleakman D, Ornstein P.L, Lodge D, Isaac J.T, Collingridge G.L. (2001) A Critical Role of a Facilitatory Presynaptic Kainate Receptor in Mossy Fiber LTP. Neuron. Nov 20;32(4):697 - 709.
28. Goff D.C, Wine L. (1997) Glutamate in schizophrenia: clinical and research implications. Schizophr Res. 27(2-3):157 - 68.
29. Ibrahim H.M, Hogg A.J Jr, Healy D.J, Haroutunian V, Davis K.L, Meador-Woodruff J.H. (2000) Ionotropic glutamate receptor binding and subunit mRNA expression in thalamic nuclei in schizophrenia. Am J Psychiatry. 157(11):1811 - 23.
30. Meador-Woodruff J.H, Davis K.L, Haroutunian V. (2001) Abnormal kainate receptor expression in prefrontal cortex in schizophrenia. Neuropsychopharmacology. 24(5):545 - 52.
31. Mohn A.R, Gainetdinov R.R, Caron M.G, Koller B.H. (1999) Mice with reduced NMDA receptor expression display behaviors related to schizophrenia. Cell. 98(4):427 - 36.
32. Sanderson T.L, Best J.J, Doody G.A, Owens D.G, Johnstone E.C (1999) Neuroanatomy of comorbid schizophrenia and learning disability: a controlled study. Lancet. 354(9193):1867 - 71.
33. Pearlson G.D. (1999) New insights on the neuroanatomy of schizophrenia. Curr Psychiatry Rep. 1(1):41 - 5.
34. Li Y, Erzurumlu R.S, Chen C, Jhaveri S, Tonegawa S (1994) Whisker-related neuronal patterns fail to develop in the trigeminal brainstem nuclei of NMDAR1 knockout mice. Cell. 76(3):427 - 37.
35. Iwasato T, Datwani A, Wolf A.M, Nishiyama H, Taguchi Y, Tonegawa S, Knopfel T, Erzurumlu R.S, Itohara S (2000) Cortex-restricted disruption of NMDAR1 impairs neuronal patterns in the barrel cortex. Nature. 406(6797):726 - 31.
36. Abel T, Lattal K.M. (2001) Molecular mechanisms of memory acquisition, consolidation and retrieval. Curr Opin Neurobiol. 11(2):180 - 7.
37. Mayford M, Abel T, Kandel E.R.(1995) Transgenic approaches to cognition. Curr Opin Neurobiol. 5(2):141 - 8.

38. Chumakov et al. (1992). Continuum of overlapping clones spanning the entire human chromosome 21q. Nature. 359, 380 - 386.

39. Osoegawa K, Mammoser A.G, Wu C, Frengen E, Zeng C, Catanese J.J, de Jong P.J. (2001) A bacterial artificial chromosome library for sequencing the complete human genome. Genome Res. 11(3):483 - 96.

40. Kozak M. (1986) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell. 44(2):283 - 92.

41. Bolger G.B, Rodgers L, Riggs M. (1994) Differential CNS expression of alternative mRNA isoforms of the mammalian genes encoding cAMP-specific phosphodiesterases. Gene. 149(2):237 - 44.

42. Huston E, Lumb S, Russell A, Catterall C, Ross A.H, Steele M.R, Bolger G.B, Perry M.J, Owens R.J, Houslay M.D. (1997) Molecular cloning and transient expression in COS7 cells of a novel human PDE4B cAMP-specific phosphodiesterase, HSPDE4B3. Biochem J. 328 (Pt 2):549 - 58.

43. Cherry J.A, Davis R.L. (1999) Cyclic AMP phosphodiesterases are localized in regions of the mouse brain associated with reinforcement, movement, and affect. J Comp Neurol. 407(2):287 - 301.

44. Barad M, Bourtchouladze R, Winder D.G, Golan H, Kandel E. (1998) Rolipram, a type IV-specific phosphodiesterase inhibitor, facilitates the establishment of long-lasting long-term potentiation and improves memory. Proc Natl Acad Sci U S A. 95(25):15020 - 5.

45. Bach M.E, Barad M, Son H, Zhuo M, Lu Y.F, Shih R, Mansuy I, Hawkins R.D, Kandel E.R. (1999) Age-related defects in spatial memory are correlated with defects in the late phase of hippocampal long-term potentiation in vitro and are attenuated by drugs that enhance the cAMP signaling pathway. Proc Natl Acad Sci U S A. 96(9):5280 - 5.

46. Gu Y.Z, Hogenesch J.B, Bradfield C.A. The PAS superfamily: sensors of environmental and developmental signals. (2000) Annu Rev Pharmacol Toxicol.;40:519 - 61.
47. Brunskill E.W, Witte D.P, Shreiner A.B, Potter S.S. (1999) Characterization of npas3, a novel basic helix-loop-helix PAS gene expressed in the developing mouse nervous system. Mech Dev. Nov;88(2):237 - 41.
48. Zhou Y.D, Barnard M, Tian H, Li X, Ring H.Z, Francke U, Shelton J, Richardson J, Russell D.W, McKnight S.L. (1997) Molecular characterization of two mammalian bHLH-PAS domain proteins selectively expressed in the central nervous system. Proc Natl Acad Sci U S A. Jan 21;94(2):713 -8.
49. Garcia J.A, Zhang D, Estill S.J, Michnoff C, Rutter J, Reick M, Scott K, Diaz-Arrastia R, McKnight S.L. (2000) Impaired cued and contextual memory in NPAS2-deficient mice. Science. Jun 23;288(5474):2226 - 30.
50. Reick M, Garcia J.A, Dudley C, McKnight S.L. (2001) NPAS2: an analog of clock operative in the mammalian forebrain. Science. Jul 20;293(5529):506 - 9.
51. Rutter J, Reick M, Wu L.C, McKnight S.L. (2001) Regulation of clock and NPAS2 DNA binding by the redox state of NAD cofactors. Science. Jul 20;293(5529):510 - 4.
52. Maemura K, Hsieh C.M, Jain M.K, Fukumoto S, Layne M.D, Liu Y, Kourembanas S, Yet S.F, Perrella M.A, Lee M.E. (1999) Generation of a dominant-negative mutant of endothelial PAS domain protein 1 by deletion of a potent C-terminal transactivation domain. J Biol Chem. Oct 29;274(44):31565 - 70.
53. Holder J.L Jr, Butte N.F, Zinn A.R. (2000) Profound obesity associated with a balanced translocation that disrupts the SIM1 gene. Hum Mol Genet. Jan 1;9(1):101 - 8.
54. Geschwind D.H, Loginov M, Stern J.M. (1999) Identification of a locus on chromosome 14q for idiopathic basal ganglia calcification (Fahr disease). Am J Hum Genet. 1999 Sep;65(3):764 - 72.

55. Wang LH, Kalb RG, Strittmatter SM. (1999) A PDZ protein regulates the distribution of the transmembrane semaphorin, M-SemF. J Biol Chem. 1999 274(20):14137-46.

56. Inagaki S, Furuyama T, Iwahashi Y. (1995) Identification of a member of mouse semaphorin family. FEBS Lett. 370(3):269-72.

57. Maziade M, Roy MA, Rouillard E, Bissonnette L, Fournier JP, Roy A, Garneau Y, Montgrain N, Potvin A, Cliche D, Dion C, Wallot H, Fournier A, Nicole L, Lavallee JC, Merette C. (2001) A search for specific and common susceptibility loci for schizophrenia and bipolar disorder: a linkage study in 13 target chromosomes. Mol Psychiatry. 6(6):684-93.

58. Hovatta I, Lichtermann D, Juvonen H, Suvisaari J, Terwilliger JD, Arajarvi R, Kokko-Sahin ML, Ekelund J, Lonnqvist J, Peltonen L. (1998) Linkage analysis of putative schizophrenia gene candidate regions on chromosomes 3p, 5q, 6p, 8p, 20p and 22q in a population-based sampled Finnish family set. Mol Psychiatry. 3(5):452-7.

59. Shimoyama Y, Tsujimoto G, Kitajima M, Natori M. (2000) Identification of three human type-II classic cadherins and frequent heterophilic interactions between different subclasses of type-II classic cadherins. Biochem J. 349(Pt 1):159-67.

60. Korematsu K, Goto S, Okamura A, Ushio Y. (1998) Heterogeneity of cadherin-8 expression in the neonatal rat striatum: comparison with striatal compartments. Exp Neurol.154(2):531-6.

61. Korematsu K, Nishi T, Okamura A, Goto S, Morioka M, Hamada J, Ushio Y. (1998) Cadherin-8 protein expression in gray matter structures and nerve fibers of the neonatal and adult mouse brain. Neuroscience. 87(1):303-15.

62. Kido M, Obata S, Tanihara H, Rochelle JM, Seldin MF, Taketani S, Suzuki ST. (1998) Molecular properties and chromosomal location of cadherin-8. Genomics. 1998 48(2):186-94.

63. Korematsu K, Redies C. (1997) Expression of cadherin-8 mRNA in the developing mouse central nervous system. J Comp Neurol.387(2):291-306.

64. Korematsu K, Redies C. (1997) Restricted expression of cadherin-8 in segmental and functional subdivisions of the embryonic mouse brain. Dev Dyn. 208(2):178-89.

65. Tanihara H, Sano K, Heimark RL, St John T, Suzuki S. (1994) Cloning of five human cadherins clarifies characteristic features of cadherin extracellular domain and provides further evidence for two structurally different types of cadherin. Cell Adhes Commun. 2(1):15-26.

66. Suzuki S, Sano K, Tanihara H. (1991) Diversity of the cadherin family: evidence for eight new cadherins in nervous tissue. Cell Regul. 2(4):261-70.

CLAIMS

1. Use of a polynucleotide fragment or fragments comprising *SEMCAP3*, *N33*, *GRIK4*, *NPAS3*, *PDE4B* and/or *CDH8* gene(s) or a fragment(s), derivative(s) or homologue(s) thereof for the manufacture of a medicament for treating schizophrenia and/or affective psychosis in a subject.

2. Use according to claim 1 wherein the SEMCAP3 nucleotide fragment comprises the sequence found in the public database under accession number AF127084 - AF127088, KIAA1095, AB029018, XM_041363 or BC014432 or the sequence shown in Figure 3.

3. Use according to either of claims 1 or 2 wherein the *N33* polynucleotide fragment comprises the sequence found in the public database under accession number U42349 or BAC RP11-23;14 or the sequences shown in Figures 6 or 7.

4. Use according to any preceding claim wherein the GRIK4 polynucleotide fragment comprises the sequence found in the public database under accession number NM_014619 or the sequences shown in Figures 10 or 16.

5. Use according to any preceding claim wherein the NPAS3 polynucleotide fragment comprises the sequence found in the public database under accession number AB054575 or AF164438 or the sequences shown in Figures 18 or 20.

6. Use according to any preceding claim wherein the PDE4B comprises the sequence as shown in Figures 25, 27 or 29.

7. Use according to any preceding claim wherein the CDH8 polynucleotide comprises the sequence found in the public database under accession number L34060, AB035305, NM_001796, AB010436, AB010437, BAC CTC-420A11 or AC040161 or as shown in Figure 35.

8. Use of a polypeptide fragment or fragments comprising *SEMCAP3*, *N33*, *GRIK4*, *NPAS3*, *PDE4B* and/or *CDH8* gene(s) or a fragment(s), derivative(s) or homologue(s) thereof for the manufacture of a medicament for treating schizophrenia and/or affective psychosis in a subject.

9. Use according to claim 8 wherein the SEMCAP3 polypeptide fragment comprises the sequence found in the public database under accession number AAF22131, AAF22132 or XP_041363, or as shown in Figure 4.

10. Use according to either of claims 8 or 9 wherein the *N33* polypeptide fragment comprises the sequence found in the public database under accession number Q13454 or as shown in Figures 6 or 7.

11. Use according to any one of claims 8 to 10 wherein the GRIK4 polypeptide fragment comprises the sequence found in the public database under accession number NM_014619, or as shown in Figures 11 and 17.

12. Use according to any one of claims 8 to 11 wherein the PDE4B polypeptide fragment comprises the sequence as shown in Figures 26, 28 or 30.

13. Use according to any one of claims 8 to 12 wherein the CDH8 polypeptide fragment comprises the sequence found in the public database under accession number NP_001787 or as shown in Figure 36.

14. Use according to any preceding claim wherein the polynucleotide fragment or polypeptide fragment consists essentially of the identified sequences.

15. A method of diagnosing schizophrenia and/or affective psychosis or susceptibility to schizophrenia and/or affective psychosis in an individual, wherein the

method comprises determining if *SEMCAP3*, *N33*, *GRIK4*, *NPAS3*, *PDE4B* and/or *CDH8* gene(s) in the individual has/have been disrupted by a mutation or chromosomal rearrangement.

16. The method according to claim 15 wherein any disruption is determined by detecting a relative level of mRNA expressed by the/said *SEMCAP3*, *N33*, *GRIK4*, *NPAS3*, *PDE4B* and/or *CDH8* gene(s).

17. The method according to claim 15 wherein a level of the/said *SEMCAP3*, *N33*, *GRIK4*, *NPAS3*, *PDE4B* and/or *CDH8* gene products are detected by an immunological technique.

18. The method according to claim 17 wherein an antibody or antibodies specific for the/said gene(s) is used to detect said gene product(s).

19. Use of an antibody or antibodies specific for SEMCAP3, *N33*, GRIK4, NPAS3, PDE4B and/or CDH8 for diagnosis of schizophrenia and/or affective psychosis.

20. Use of an antibody or antibodies specific for SEMCAP3, *N33*, GRIK4, NPAS2, PDE4B and/or CDH8 for the manufacture of a medicament for the treatment of schizophrenia and/or affective psychosis.

21. An animal model for psychiatric disorders wherein the animal model has been generated by specifically disruption expression of the/said *SEMCAP3*, *N33*, *GRIK4*, *NPAS2*, *PDE4B* and/or *CDH8* gene(s).

22. An animal model for psychiatric disorder wherein the animal model has been generated by specifically upregulating expression of the/said *SEMCAP3*, *N33*, *GRIK4*, *NPAS2*, *PDE4B* and/or *CDH8* gene(s).

23. A method for identifying ligands for *SEMCAP3*, *N33*, *GRIK4*, *NPAS2*, *PDE4B* and/or *CDH8* gene(s) products, said method comprising the steps of:

a) introducing into a suitable host cell a polynucleotide fragment according to the invention;

b) culturing cells under conditions to allow expression of the polynucleotide fragment;

c) optionally isolating the expression product;

d) brining the expression product (or the host cell from step b)) into contact with potential ligands which will possibly bind to the protein encoded by said polynucleotide fragment from step a);

e) establishing whether a ligand has bound to the expressed protein; and

f) optionally isolating and identifying the ligand.

Figure 1

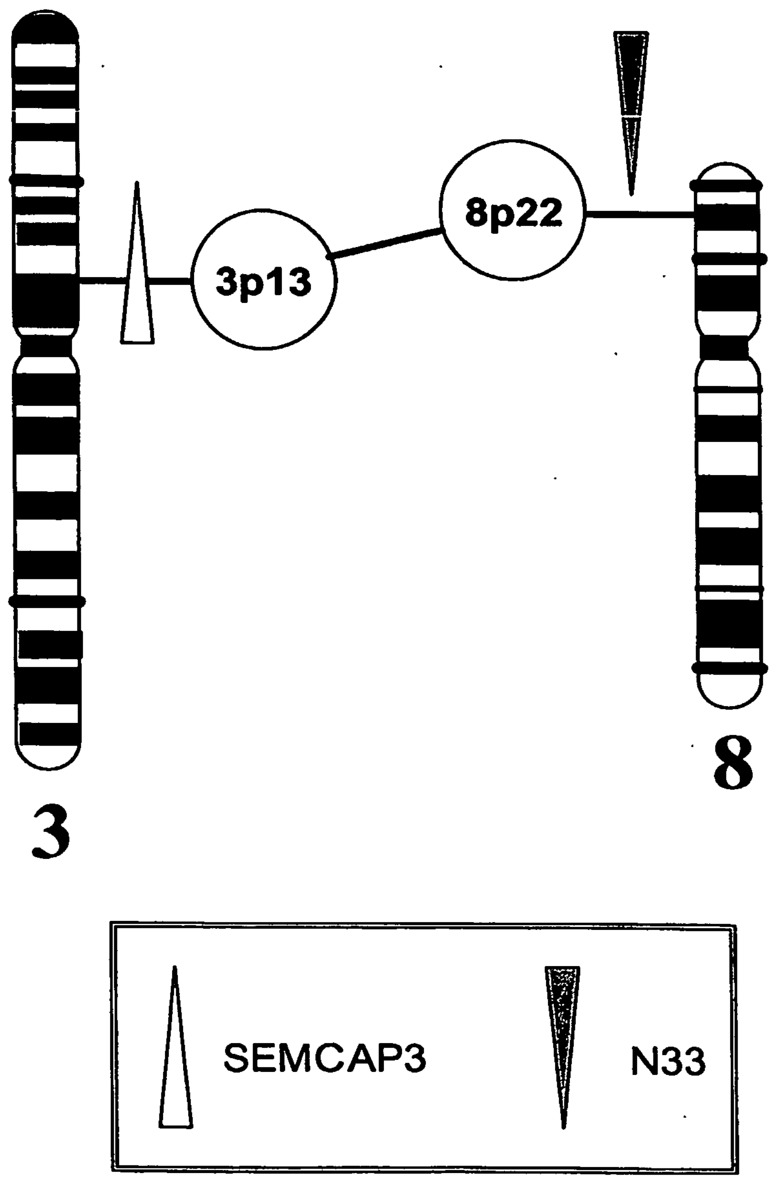


Figure 2

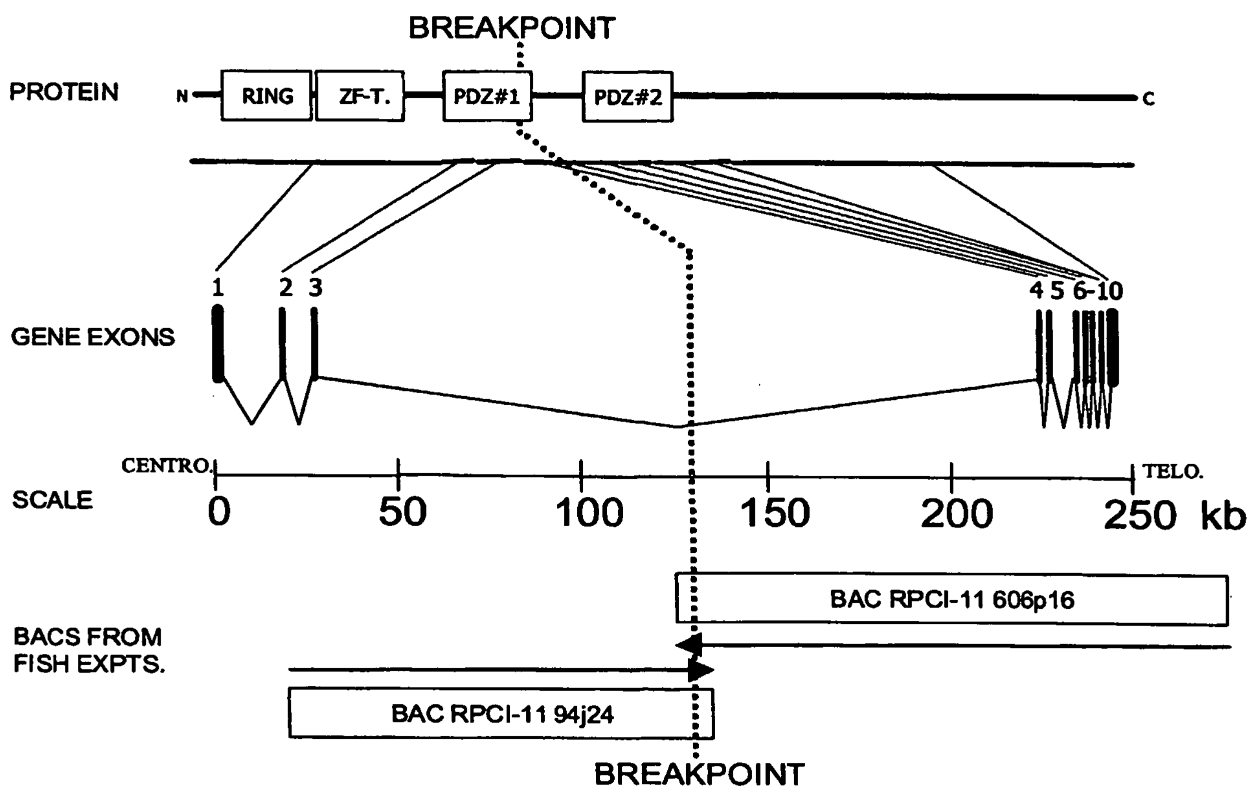

BREAKPOINT
PROTEIN
N
RING
ZF-T.
PDZ#1
PDZ#2
C
1
2
3
4 5 6-10
GENE EXONS
CENTRO.
TELO.
SCALE
0
50
100
150
200
250 kb
BAC RPCI-11 606p16
BACS FROM FISH EXPTS.
BAC RPCI-11 94j24
BREAKPOINT

Figure 3

```
   1 AAAACTTCCC CGGGTAGATT CACCCACCGG TCCTGGAAAC CTGCTAAATC CTGAAGGTTC
  61 ACAGAACCTC TGGTCAGAAC TGAAGTTGCA GCCGGAGCTT CCCGCAGGCT CTGTAACTTT
 121 CCCTGGAATG AAATAAATAA ATAAAGACCG TAAGTGCTGA GATAGCGGGC CCCAAGATAT
 181 TTTTAGTCCT CTGCAATCAG CCACTAGAGG AAGGGGGAGG GAGAAGGGAG TAAAAAAGTT
 241 TTGATCCGTT CGGGAAGGGG CTCGAAGAGA ACCCTTGGGA GAAAGCAGTA GCCTCAGCTC
 301 CAAACTCAGC GAGCTTTTCT CGGCTGGCGT TTTGTCTCCT ATAGCGTAGA CTGTAAGAGA
 361 ACAGAAAGGA GTTTCCCGAG AAGATTCAGG CTGGCGTCCT GGGCTGGCCC GTCCCTTCTG
 421 GCGAGCCTCA GTGTCCTCCC ACGCGCTTCT GCCTTCCAGC CTCCTCCCTT TTTCGGGGGG
 481 CTGGCGGGAG GCATCCAAGG CACGATGTAT GTGCGCTCGC GCTCGCGCAA ATACGGCCGG
 541 AGGAGTCCTG TTCCTCGGGC ATTTTCCGAG GAAGTCTGGA TCAATTAGGC TCAGTCCGGG
 601 GAGAGCCAGC GAGCGCGCGG GCGGCGTAGC CGGCCTGTCT GGGCCGCCTC GTGGGGAGGG
 661 AGGGGGCGCC CGGCCGCCCG GCGGCGACCC CGGGGCCTGG CCGCCACCAT GGGCTTCGAG
 721 CTGGACCGCT TCGACGGCGA CGTGGACCCG GACCTGAAGT GCGCGCTGTG CCACAAGGTC
 781 CTGGAGGACC CGCTGACCAC GCCGTGCGGC CACGTCTTCT GCGCCGGCTG CGTGCTGCCC
 841 TGGGTGGTGC AGGAGGGCAG CTGCCCGGCG CGCTGCCGCG GTCGCCTGTC GGCCAAAGAG
 901 CTCAACCACG TCCTGCCGCT CAAGCGCCTT ATCCTCAAGC TGGACATCAA GTGCGCGTAC
 961 GCGACGCGCG GCTGCGGCCG GGTGGTCAAG CTGCAGCAGC TGCCGGAGCA CCTCGAGCGC
1021 TGCGACTTCG CGCCCGCGCG CTGTCGCCAC GCGGGTTGCG GCCAGGTGCT GCTGCGGCGC
1081 GACGTGGAGG CGCACATGCG CGACGCGTGC GACGCGCGGC CAGTGGGCCG CTGCCAGGAG
1141 GGCTGCGGGC TACCCTTGAC GCACGGCGAG CAGCGCGCGG GCGGCCACTG CTGCGCGCGA
1201 GCGCTGCGGG CGCACAACGG CGCGCTCCAG GCCCGCCTGG GCGCGCTGCA CAAGGCGCTC
1261 AAGAAGGAGG CGCTGCGCGC TGGGAAGCGC GAGAAGTCGC TGGTGGCCCA GCTGGCCGCG
1321 GCGCAGCTTG AGCTGCAGAT GACCGCGCTG CGCTACCAGA AGAAATTCAC CGAATACAGC
1381 GCGCGCCTCG ACTCGCTCAG CCGCTGCGTG GCCGCGCCGC CCGGCGGCAA GGGCGAAGAA
1441 ACCAAAAGTC TGACTCTTGT CCTGCATCGG GACTCCGGCT CCCTGGGATT CAATATTATT
1501 GGTGGCCGGC CGAGTGTGGA TAACCACGAT GGATCATCCA GTGAAGGAAT CTTTGTATCC
1561 AAGATAGTTG ACAGTGGGCC TGCAGCCAAG GAAGGAGGCC TGCAAATTCA TGACAGGATT
1621 ATTGAGGTCA ACGGCAGAGA CTTATCCAGA GCAACTCATG ACCAGGCTGT GGAAGCTTTC
1681 AAGACAGCCA AGGAGCCCAT AGTGGTGCAG GTGTTGAGAA GAACACCAAG GACCAAAATG
1741 TTCACGCCTC CATCAGAGTC TCAGCTGGTG GACACGGGAA CCCAAACCGA CATCACCTTT
1801 GAACATATCA TGGCCCTCAC TAAGATGTCC TCTCCCAGCC CACCCGTGCT GGATCCCTAT
1861 CTCTTGCCAG AGGAGCATCC CTCAGCCCAT GAATACTACG ATCCAAATGA CTACATTGGA
1921 GACATCCATC AGGAGATGGA CAGGGAGGAG CTGGAGCTGG AGGAAGTGGA CCTCTACAGA
1981 ATGAACAGCC AGGACAAGCT GGGCCTCACT GTGTGCTACC GGACGGACGA TGAAGACGAC
```

2041 ATTGGGATTT ATATCAGTGA GATTGACCCT AACAGCATTG CAGCCAAGGA TGGGCGCATC
2101 CGAGAAGGAG ACCGCATTAT CCAGATTAAT GGGATAGAGG TGCAGAACCG TGAAGAGGCT
2161 GTGGCTCTTC TAACCAGTGA AGAAAATAAA AACTTTTCAT TGCTGATTGC AAGGCCTGAA
2221 CTCCAGCTGG ATGAGGGCTG GATGGATGAT GACAGGAACG ACTTTCTGGA TGACCTGCAC
2281 ATGGACATGC TGGAGGAGCA GCACCACCAG GCCATGCAAT TCACAGCTAG CGTGCTGCAG
2341 CAGAAGAAGC ACGACGAAGA CGGTGGGACC ACAGATACAG CCACCATCTT GTCCAACCAG
2401 CACGAGAAGG ACAGCGGTGT GGGGCGGACC GACGAGAGCA CCCGTAATGA CGAGAGCTCG
2461 GAGCAAGAGA ACAATGGCGA CGACGCCACC GCATCCTCCA ACCCGCTGGC GGGGCAGAGG
2521 AAGCTCACCT GCAGCCAGGA CACCTTGGGC AGCGGCGACC TGCCCTTCAG CAACGAGTCT
2581 TTCATTTCGG CCGACTGCAC GGACGCCGAC TACCTGGGGA TCCCGGTGGA CGAGTGCGAG
2641 CGCTTCCGCG AGCTCCTGGA GCTCAAGTGC CAGGTGAAGA GCGCCACCCC TTACGGCCTG
2701 TACTACCCTA GCGGCCCCCT GGACGCCGGC AAGAGTGACC CTGAGAGCGT GGACAAGGAG
2761 CTGGAGCTGC TGAACGAAGA GCTGCGCAGC ATCGAGCTGG AGTGCCTGAG CATCGTGCGC
2821 GCCCACAAGA TGCAGCAGCT CAAGGAGCAG TACCGCGAGT CCTGGATGCT GCACAACAGC
2881 GGCTTCCGCA ACTACAACAC CAGCATCGAC GTGCGCAGAC ACGAGCTCTC AGATATCACC
2941 GAGCTCCCGG AGAAATCCGA CAAGGACAGC TCGAGCGCCT ACAACACAGG CGAGAGCTGC
3001 CGCAGCACCC CGCTCACCCT GGAGATCTCC CCCGACAACT CCTTGAGGAG AGCGGCGGAG
3061 GGCATCAGCT GCCCGAGCAG CGAAGGGGCT GTGGGGACCA CGGAAGCCTA CGGGCCAGCC
3121 TCCAAGAATC TGCTCTCCAT CACGGAAGAT CCCGAAGTGG GCACCCCTAC CTATAGCCCG
3181 TCCCTGAAGG AGCTGGACCC CAACCAGCCC CTGGAAAGCA AAGAGCGGAG AGCCAGCGAC
3241 GGGAGCCGGA GCCCCACGCC CAGCCAGAAG CTGGGCAGCG CCTACCTGCC CTCCTATCAC
3301 CACTCCCCAT ACAAGCACGC GCACATCCCG GCGCACGCCC AGCACTACCA GAGCTACATG
3361 CAGCTGATCC AGCAGAAGTC GGCCGTGGAG TACGCGCAAA GCCAGATGAG CCTGGTGAGC
3421 ATGTGCAAGG ACCTGAGCTC TCCCACCCCG TCGGAGCCGC GCATGGAGTG GAAGGTGAAG
3481 ATCCGCAGCG ACGGGACGCG CTACATCACC AAGAGGCCCG TGCGGGACCG CCTGCTGCGG
3541 GAGCGCGCCC TGAAGATCCG GGAAGAGCGC AGCGGCATGA CCACCGACGA CGACGCGGTG
3601 AGCGAGATGA AGATGGGGCG CTACTGGAGC AAGGAGGAGA GGAAGCAGCA CCTGGTGAAG
3661 GCCAAGGAGC AGCGGCGGCG GCGCGAGTTC ATGATGCAGA GCAGGTTGGA TTGTCTCAAG
3721 GAGCAGCAAG CAGCCGATGA CAGGAAGGAG ATGAACATTC TCGAACTGAG CCACAAAAAG
3781 ATGATGAAGA AGAGGAATAA GAAAATCTTC GATAACTGGA TGACGATCCA AGAACTCTTA
3841 ACCCACGGCA CAAAATCCCC GGACGGCACT AGAGTATACA ATTCCTTCCT ATCGGTGACT
3901 ACTGTA<u>TAA</u>T TTTCACTTCT GCATTATGTA CATAAAGGAG ACCACTACCA CTGGGGTAGA
3961 AATTCCTGCC TCGTTCAATG CGGCAAGTTT TTGTATATAA GATAAGTACG GTCTTCATGT
4021 TTATAGTCCA AATTTGCAAA CCCTACAACT CTGGGTGTCA TAGGTCTATT TTAAGGGAAG
4081 AGAGAGAAAA ACACCCTTAC TATCTTGGAA GGCAATATTA ACAAACAGAG CTTTTTTCAA
4141 ATAGCAATTG TACTTTTCTA CCTGTACCCT TTTACATAAA GTGTTTAAAT TTCAGAAAGA

4201 TCTTTTATTA AGCATACTTT CACAGAATAA CTTGTTTAAA CTATATTCAT ATAAAAAAGT
4261 TAAACACGCT TTTTTTCCTG CCTAAAACAC AAATACAACT GCCAGTATGT ATTTTTAATG
4321 GAACCCTATT TTATAATGGT ACGTTACTGA ATGTGTTTCA TATGCGTGAC CGTTAAGATA
4381 TTATCATTTA GGTGAAGGTT TCAACTCAAA ACCACCCAAC CCGGTGGTTA ACGATTTAAT
4441 ACATATAACC AAACCGGCAG CGTTTAGAGT TGGGATATAC ATTTAAACAT TTTCCTGGTT
4501 AAGGTTCCCA AGAGAGTGTA AAGGTTTTAG CAGAAAGCAA AATATCTTGC ATCTTTATGG
4561 AAGTTTAAAG CATGTTTGCA AATATTGCAG CCCATTGAAA GAATTTGCAT GTACAGGAAA
4621 GTTGTGGATG GAGACGGTTT GTGGAATTTT AAGTGCTCAT TGTAGTAAAC TTTTGCTTTG
4681 TAGATTTGAA GGTACAGACT TATACAGGCA AGTTCACAAA ATCATGATTA GTTACAAACA
4741 GTAAAATGAA GTTAAAATAA ATTATTATTT TCT

Figure 4

```
1 MGFELDRFDG DVDPDLKCAL CHKVLEDPLT TPCGHVFCAG CVLPWVVQEG SCPARCRGRL
61 SAKELNHVLP LKRLILKLDI KCAYATRGCG RVVKLQQLPE HLERCDFAPA RCRHAGCGQV
121 LLRRDVEAHM RDACDARPVG RCQEGCGLPL THGEQRAGGH CCARALRAHN GALQARLGAL
181 HKALKKEALR AGKREKSLVA QLAAAQLELQ MTALRYQKKF TEYSARLDSL SRCVAAPPGG
241 KGEETKSLTL VLHRDSGSLG FNIIGGRPSV DNHDGSSSEG IFVSKIVDSG PAAKEGGLQI
301 HDRIIEVNGR DLSRATHDQA VEAFKTAKEP IVVQVLRRTP RTKMFTPPSE SQLVDTGTQT
361 DITFEHIMAL TKMSSPSPPV LDPYLLPEEH PSAHEYYDPN DYIGDIHQEM DREELELEEV
421 DLYRMNSQDK LGLTVCYRTD DEDDIGIYIS EIDPNSIAAK DGRIREGDRI IQINGIEVQN
481 REEAVALLTS EENKNFSLLI ARPELQLDEG WMDDDRNDFL DDLHMDMLEE QHHQAMQFTA
541 SVLQQKKHDE DGGTTDTATI LSNQHEKDSG VGRTDESTRN DESSEQENNG DDATASSNPL
601 AGQRKLTCSQ DTLGSGDLPF SNESFISADC TDADYLGIPV DECERFRELL ELKCQVKSAT
661 PYGLYYPSGP LDAGKSDPES VDKELELLNE ELRSIELECL SIVRAHKMQQ LKEQYRESWM
721 LHNSGFRNYN TSIDVRRHEL SDITELPEKS DKDSSSAYNT GESCRSTPLT LEISPDNSLR
781 RAAEGISCPS SEGAVGTTEA YGPASKNLLS ITEDPEVGTP TYSPSLKELD PNQPLESKER
841 RASDGSRSPT PSQKLGSAYL PSYHHSPYKH AHIPAHAQHY QSYMQLIQQK SAVEYAQSQM
901 SLVSMCKDLS SPTPSEPRME WKVKIRSDGT RYITKRPVRD RLLRERALKI REERSGMTTD
961 DDAVSEMKMG RYWSKEERKQ HLVKAKEQRR RREFMMQSRL DCLKEQQAAD DRKEMNILEL
1021 SHKKMMKKRN KKIFDNWMTI QELLTHGTKS PDGTRVYNSF LSVTTV
```

Figure 5

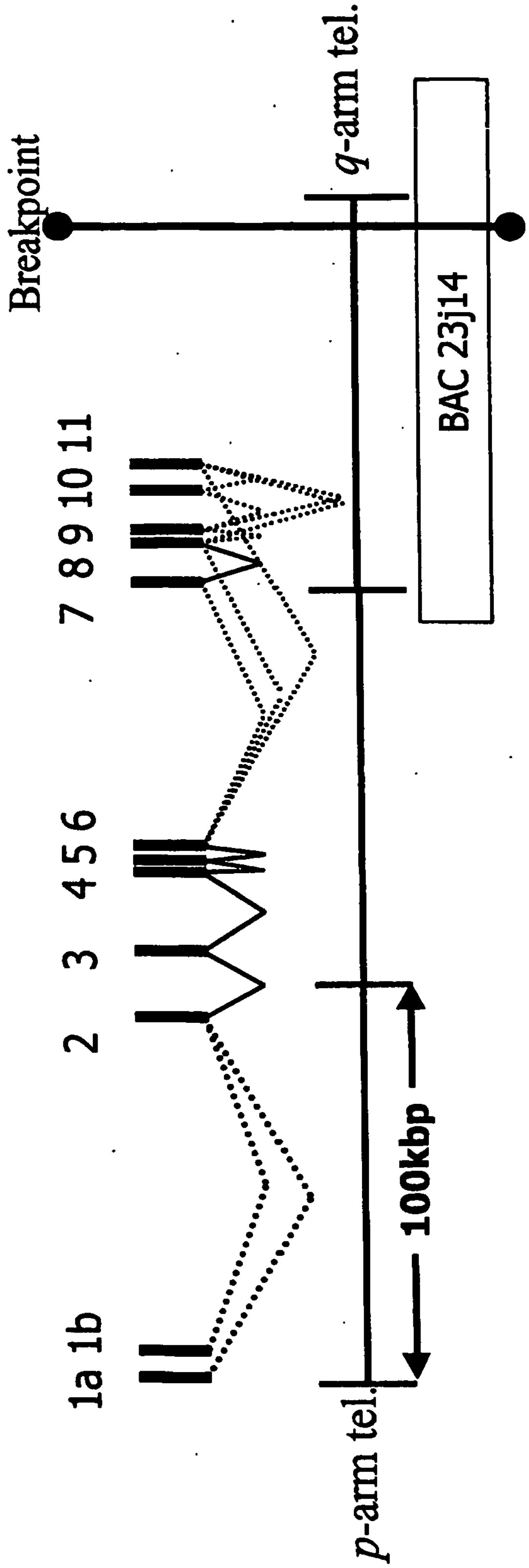

Breakpoint
q-arm tel.
BAC 23j14
7 8 9 10 11
4 5 6
3
2
1a 1b
p-arm tel.
100kbp

Figure 6

1a
atcttcctcctgctctggctgtgtgaagatctgcctccttcctcttcggcttcatgcat
gatcgtaagtttcctgaggcctcctcagccatgcttcctgcatagcctgcagaaat

1b
cccgggtccctcgcaaagccgctgccatcccggagggcccagccagcgggctcccggag
gctggccgggcaggcgtggtgcgcggtaggagctgggcgcgcacggctaccgcgcgtgg
aggagacactgccctgccgcgatgggggcccggggcgctccttcacgccgtaggcaagc
ggggcggcggctgcggtacctgcccaccgggagctttcccttccttctcctgctgctgc
tgctctgcatccagctcgggggaggacagaagaaaaaggag

2-6
These exons have been joined together as they are always spliced in this way.
aatcttttagctgaaaaagtagagcagctgatggaatggagttccagacgctcaatctt
ccgaatgaatggtgataaattccgaaaatttataaaggcaccacctcgaaactattcca
tgattgttatgttcactgctcttcagcctcagcggcagtgttctgtgtgcaggcaagct
aatgaagaatatcaaatactggcgaactcctggcgctattcatctgctttttgtaacaa
gctcttcttcagtatggtggactatgatgaggggacagacgttttcagcagctcaaca
tgaactctgctcctacattcatgcatttcctccaaaaggcagacctaagagagctgat
acttttgacctccaaagaattggatttgcagctgagcaactagcaaagtggattgctga
cagaacggatgttcatattcgggttttcagaccacccaactactctggtaccattgctt
tggccctgttagtgtcgcttgttggaggtttgctttatttgagaaggaacaacttggag
ttcatctataacaagactggttgggccatggtgtctctgtgtatagtctttgctatgac
ttctggccagatgtggaaccatatccgtggacctccatatgctcataagaacccacaca
atggacaagtg

7
agctacattcatgggagcagccaggctcagtttgtggcagaatcacacattattctggt
actga

8
atgccgctatcaccatggggatggttcttctaaatgaagcagcaacttcgaaaggcgat
gttggaaaaagacgga

8+
This is identical to 8 except a cryptic splice acceptor upstream is employed.
tttaaccattctggaacattgtgttcagagccagaaaaattaatagatttattcacat
ctatgtctacggcttccttgacaactactgcagatgccgctatcaccatggggatggtt
cttctaaatgaagcagcaacttcgaaaggcgatgttggaaaaagacgga

9
taatttgcctagtgggattgggcctggtggtcttcttcttcagttttctactttcaata
tttcgttccaagtaccacggctatccttatag

10
tgatctggactttgagtgagaagatgtgatttggaccatggcacttaaaaactctataa
cctcag

11
ctttttaattaaatgaagccaagtgggatttgcataaagtgaatgtttaccatgaagat
aaactgttcctgactttatactattttgaattc

Figure 7

Alternative start exons

1a:

MEWSSRRSIFRMNGDKFRKFIKAPPRNYS………… (encoded by exon 2).

1b:

MGARGAPSRRRQAGRRLRYLPTGSFPFLLLLLLLCIQLGGGQKKKENLLAEKVEQLMEW
SSRRSIFRMNGDKFRKFIKAPPRNYS…………

Transcript options

2-6,7,8,9,10,11
aatcttttagctgaaaagtagagcagctgatggaatggagttccagacgctcaatctt
ccgaatgaatggtgataaattccgaaaatttataaaggcaccacctcgaaactattcca
tgattgttatgttcactgctcttcagcctcagcggcagtgttctgtgtgcaggcaagct
aatgaagaatatcaaatactggcgaactcctggcgctattcatctgcttttgtaacaa
gctcttcttcagtatggtggactatgatgaggggacagacgtttttcagcagctcaaca
tgaactctgctcctacattcatgcatttcctccaaaaggcagacctaagagagctgat
acttttgacctccaaagaattggatttgcagctgagcaactagcaaagtggattgctga
cagaacggatgttcatattcgggttttcagaccacccaactactctggtaccattgctt
tggccctgttagtgtcgcttgttggaggtttgctttatttgagaaggaacaacttggag
ttcatctataacaagactggttgggccatggtgtctctgtgtatagtctttgctatgac
ttctggccagatgtggaaccatatccgtggacctccatatgctcataagaacccacaca
atggacaagtgagctacattcatgggagcagccaggctcagtttgtggcagaatcacac
attattctggtactgaatgccgctatcaccatggggatggttcttctaaatgaagcagc
aacttcgaaaggcgatgttggaaaaagacggataatttgcctagtgggattgggcctgg
tggtcttcttcttcagttttctactttcaatatttcgttccaagtaccacggctatcct
tatagtgatctggactttgagtgagaagatgtgatttggaccatggcacttaaaaactc
tataacctcagctttttaattaaatgaagccaagtgggatttgcataaagtgaatgttt
accatgaagataaactgttcctgactttatactattttgaattc

(MGARGAPSRRRQAGRRLRYLPTGSFPFLLLLLLLCIQLGGGQKKKENLLAEKVEQL)M
EWSSRRSIFRMNGDKFRKFIKAPPRNYSMIVMFTALQPQRQCSVCRQANEEYQILANSW
RYSSAFCNKLFFSMVDYDEGTDVFQQLNMNSAPTFMHFPPKGRPKRADTFDLQRIGFAA
EQLAKWIADRTDVHIRVFRPPNYSGTIALALLVSLVGGLLYLRRNNLEFIYNKTGWAMV
SLCIVFAMTSGQMWNHIRGPPYAHKNPHNGQVSYIHGSSQAQFVAESHIILVLNAAITM
GMVLLNEAATSKGDVGKRRIICLVGLGLVVFFFSFLLSIFRSKYHGYPYSDLDFE

2-6,7,8,9,11

aatcttttagctgaaaaagtagagcagctgatggaatggagttccagacgctcaatctt
ccgaatgaatggtgataaattccgaaaatttataaaggcaccacctcgaaactattcca
tgattgttatgttcactgctcttcagcctcagcggcagtgttctgtgtgcaggcaagct
aatgaagaatatcaaatactggcgaactcctggcgctattcatctgcttttgtaacaa
gctcttcttcagtatggtggactatgatgaggggacagacgttttcagcagctcaaca
tgaactctgctcctacattcatgcattttcctccaaaaggcagacctaagagagctgat
acttttgacctccaaagaattggatttgcagctgagcaactagcaaagtggattgctga
cagaacggatgttcatattcgggttttcagaccacccaactactctggtaccattgctt
tggccctgttagtgtcgcttgttggaggtttgctttatttgagaaggaacaacttggag
ttcatctataacaagactggttgggccatggtgtctctgtgtatagtctttgctatgac
ttctggccagatgtggaaccatatccgtggacctccatatgctcataagaacccacaca
atggacaagtgagctacattcatgggagcagccaggctcagtttgtggcagaatcacac
attattctggtactgaatgccgctatcaccatggggatggttcttctaaatgaagcagc
aacttcgaaaggcgatgttggaaaaagacggataatttgcctagtgggattggcctgg
tggtcttcttcttcagttttctactttcaatatttcgttccaagtaccacggctatcct
tatagcttttaattaaatgaagccaagtgggatttgcataaagtgaatgtttaccatg
aagataaactgttcctgactttatactattttgaattc

(MGARGAPSRRRQAGRRLRYLPTGSFPFLLLLLLLCIQLGGGQKKKENLLAEKVEQL)M
EWSSRRSIFRMNGDKFRKFIKAPPRNYSMIVMFTALQPQRQCSVCRQANEEYQILANSW
RYSSAFCNKLFFSMVDYDEGTDVFQQLNMNSAPTFMHFPPKGRPKRADTFDLQRIGFAA
EQLAKWIADRTDVHIRVFRPPNYSGTIALALLVSLVGGLLYLRRNNLEFIYNKTGWAMV
SLCIVFAMTSGQMWNHIRGPPYAHKNPHNGQVSYIHGSSQAQFVAESHIILVLNAAITM
GMVLLNEAATSKGDVGKRRIICLVGLGLVVFFFSFLLSIFRSKYHGYPYSFLIK

2-6,11
aatcttttagctgaaaaagtagagcagctgatggaatggagttccagacgctcaatctt
ccgaatgaatggtgataaattccgaaaatttataaaggcaccacctcgaaactattcca
tgattgttatgttcactgctcttcagcctcagcggcagtgttctgtgtgcaggcaagct
aatgaagaatatcaaatactggcgaactcctggcgctattcatctgcttttgtaacaa
gctcttcttcagtatggtggactatgatgaggggacagacgttttcagcagctcaaca
tgaactctgctcctacattcatgcattttcctccaaaaggcagacctaagagagctgat
acttttgacctccaaagaattggatttgcagctgagcaactagcaaagtggattgctga
cagaacggatgttcatattcgggttttcagaccacccaactactctggtaccattgctt
tggccctgttagtgtcgcttgttggaggtttgctttatttgagaaggaacaacttggag
ttcatctataacaagactggttgggccatggtgtctctgtgtatagtctttgctatgac
ttctggccagatgtggaaccatatccgtggacctccatatgctcataagaacccacaca
atggacaagtgcttttaattaaatgaagccaagtgggatttgcataaagtgaatgttt
accatgaagataaactgttcctgactttatactattttgaattc

(MGARGAPSRRRQAGRRLRYLPTGSFPFLLLLLLLCIQLGGGQKKKENLLAEKVEQL)M
EWSSRRSIFRMNGDKFRKFIKAPPRNYSMIVMFTALQPQRQCSVCRQANEEYQILANSW
RYSSAFCNKLFFSMVDYDEGTDVFQQLNMNSAPTFMHFPPKGRPKRADTFDLQRIGFAA

EQLAKWIADRTDVHIRVFRPPNYSGTIALALLVSLVGGLLYLRRNNLEFIYNKTGWAMV
SLCIVFAMTSGQMWNHIRGPPYAHKNPHNGQVLFN

2-6,7,8,11
aatcttttagctgaaaaagtagagcagctgatggaatggagttccagacgctcaatctt
ccgaatgaatggtgataaattccgaaaatttataaaggcaccacctcgaaactattcca
tgattgttatgttcactgctcttcagcctcagcggcagtgttctgtgtgcaggcaagct
aatgaagaatatcaaatactggcgaactcctggcgctattcatctgcttttgtaacaa
gctcttcttcagtatggtggactatgatgaggggacagacgttttcagcagctcaaca
tgaactctgctcctacattcatgcatttcctccaaaaggcagacctaagagagctgat
acttttgacctccaaagaattggatttgcagctgagcaactagcaaagtggattgctga
cagaacggatgttcatattcgggttttcagaccacccaactactctggtaccattgctt
tggccctgttagtgtcgcttgttggaggtttgctttatttgagaaggaacaacttggag
ttcatctataacaagactggttgggccatggtgtctctgtgtatagtctttgctatgac
ttctggccagatgtggaaccatatccgtggacctccatatgctcataagaacccacaca
atggacaagtgagctacattcatgggagcagccaggctcagtttgtggcagaatcacac
attattctggtactgaatgccgctatcaccatggggatggttcttctaaatgaagcagc
aacttcgaaaggcgatgttggaaaaagacggacttttaattaaatgaagccaagtggg
atttgcataaagtgaatgtttaccatgaagataaactgttcctgactttatactatttt
gaattc

(MGARGAPSRRRQAGRRLRYLPTGSFPFLLLLLLLCIQLGGGQKKKENLLAEKVEQL)M
EWSSRRSIFRMNGDKFRKFIKAPPRNYSMIVMFTALQPQRQCSVCRQANEEYQILANSW
RYSSAFCNKLFFSMVDYDEGTDVFQQLNMNSAPTFMHFPPKGRPKRADTFDLQRIGFAA
EQLAKWIADRTDVHIRVFRPPNYSGTIALALLVSLVGGLLYLRRNNLEFIYNKTGWAMV
SLCIVFAMTSGQMWNHIRGPPYAHKNPHNGQVSYIHGSSQAQFVAESHIILVLNAAITM
GMVLLNEAATSKGDVGKRRTF

2-6,8+,9,11
aatcttttagctgaaaaagtagagcagctgatggaatggagttccagacgctcaatctt
ccgaatgaatggtgataaattccgaaaatttataaaggcaccacctcgaaactattcca
tgattgttatgttcactgctcttcagcctcagcggcagtgttctgtgtgcaggcaagct
aatgaagaatatcaaatactggcgaactcctggcgctattcatctgcttttgtaacaa
gctcttcttcagtatggtggactatgatgaggggacagacgttttcagcagctcaaca
tgaactctgctcctacattcatgcatttcctccaaaaggcagacctaagagagctgat
acttttgacctccaaagaattggatttgcagctgagcaactagcaaagtggattgctga
cagaacggatgttcatattcgggttttcagaccacccaactactctggtaccattgctt
tggccctgttagtgtcgcttgttggaggtttgctttatttgagaaggaacaacttggag
ttcatctataacaagactggttgggccatggtgtctctgtgtatagtctttgctatgac
ttctggccagatgtggaaccatatccgtggacctccatatgctcataagaacccacaca
atggacaagtgtttaaccattctggaacattgtgttcagagccagaaaaattaatagat
tttattcacatctatgtctacggcttccttgacaactactgcagatgccgctatcacca
tggggatggttcttctaaatgaagcagcaacttcgaaaggcgatgttggaaaaagacgg
ataatttgcctagtgggattgggcctggtggtcttcttcttcagttttctactttcaat
atttcgttccaagtaccacggctatccttatagctttttaattaaatgaagccaagtgg

gatttgcataaagtgaatgtttaccatgaagataaactgttcctgactttatactattt
tgaattc

(MGARGAPSRRRQAGRRLRYLPTGSFPFLLLLLLLCIQLGGGQKKKENLLAEKVEQL)M
EWSSRRSIFRMNGDKFRKFIKAPPRNYSMIVMFTALQPQRQCSVCRQANEEYQILANSW
RYSSAFCNKLFFSMVDYDEGTDVFQQLNMNSAPTFMHFPPKGRPKRADTFDLQRIGFAA
EQLAKWIADRTDVHIRVFRPPNYSGTIALALLVSLVGGLLYLRRNNLEFIYNKTGWAMV
SLCIVFAMTSGQMWNHIRGPPYAHKNPHNGQVFNHSGTLCSEPEKLIDFIHIYVYGFLD
NYCRCRYHHGDGSSK

2-6,8+,11
aatcttttagctgaaaaagtagagcagctgatggaatggagttccagacgctcaatctt
ccgaatgaatggtgataaattccgaaaatttataaaggcaccacctcgaaactattcca
tgattgttatgttcactgctcttcagcctcagcggcagtgttctgtgtgcaggcaagct
aatgaagaatatcaaatactggcgaactcctggcgctattcatctgcttttttgtaacaa
gctcttcttcagtatggtggactatgatgaggggacagacgtttttcagcagctcaaca
tgaactctgctcctacattcatgcatttcctccaaaaggcagacctaagagagctgat
acttttgacctccaaagaattggatttgcagctgagcaactagcaaagtggattgctga
cagaacggatgttcatattcgggttttcagaccacccaactactctggtaccattgctt
tggccctgttagtgtcgcttgttggaggtttgctttatttgagaaggaacaacttggag
ttcatctataacaagactggttgggccatggtgtctctgtgtatagtctttgctatgac
ttctggccagatgtggaaccatatccgtggacctccatatgctcataagaacccacaca
atggacaagtgtttaaccattctggaacattgtgttcagagccagaaaaattaatagat
tttattcacatctatgtctacggcttccttgacaactactgcagatgccgctatcacca
tggggatggttcttctaaatgaagcagcaacttcgaaaggcgatgttggaaaaagacgg
acttttaattaaatgaagccaagtgggatttgcataaagtgaatgtttaccatgaaga
taaactgttcctgactttatactattttgaattc

(MGARGAPSRRRQAGRRLRYLPTGSFPFLLLLLLLCIQLGGGQKKKENLLAEKVEQL)M
EWSSRRSIFRMNGDKFRKFIKAPPRNYSMIVMFTALQPQRQCSVCRQANEEYQILANSW
RYSSAFCNKLFFSMVDYDEGTDVFQQLNMNSAPTFMHFPPKGRPKRADTFDLQRIGFAA
EQLAKWIADRTDVHIRVFRPPNYSGTIALALLVSLVGGLLYLRRNNLEFIYNKTGWAMV
SLCIVFAMTSGQMWNHIRGPPYAHKNPHNGQVFNHSGTLCSEPEKLIDFIHIYVYGFLD
NYCRCRYHHGDGSSK

Figure 8

```
IAG2_HUMAN            MAAR------------WRFWCVSVTMVVALLIVCDVPSASA
N33_HUMAN             MGARGAPSRRRQAGRRLRYLPTGSFPFLLLLLLLCIQLGGG
DROS._CG7830          ----------------MRLLHKTLLSGLLVVALFAIYAAAQ
Celegans_g304348      -------------------------------MLLAVYESAQ
Yeast_Ost3p           --------------------------MNWLFLVSLVFFCGV
Yeast_Ost6p           ----------------------MKWCSTYIIIWLAIIFHKF

IAG2_HUMAN            QRKKE-MVLSEKVSQLMEWTNKRPVIRMNGDKFRRLVKAPF
N33_HUMAN             QKKKE-NLLAEKVEQLMEWSSRRSIFRMNGDKFRKFIKAPF
DROS._CG7830          SKSKTGLSLSEKVQNLVDMNAKKPLLRFNGPKFREYVKSAF
Celegans_g304348      QQT-----LEDKVQNLVDLTSRQSIVKFNMDKWKTLVRMQF
Yeast_Ost3p           STHPALAMSSNRLLKLANKSPKK---IIPLKDSSFENILAF
Yeast_Ost6p           QKSTA--TASHNIDDILQLKDDTGVITVTADNYPLLSRGVF

IAG2_HUMAN            --RNYSVLVMFPALQLHRQCVVCKQADEEFQILANSWRYSS
N33_HUMAN             --RNYSMIVMFPALQPQRQCSVCRQANEEYQILANSWRYSS
DROS._CG7830          --RNYSMIVMLPALAPSRQCQICRHAHDEFAIVANSYRFSS
Celegans_g304348      --RNYSMIVMFPALSPGVQCPICKPAYDEFMIVANSHRYTS
Yeast_Ost3p           PHENAYIVALFTATAPEIGCSLCLELESEYDTIVASWFDDH
Yeast_Ost6p           GYFNILYITMRGTNSNGMSCQLCHDFEKTYHAVEDVIKSQA
                                         CYST.

IAG2_HUMAN            AFTN------RIEFAMVDFDEG----SDVFQMLNMNSAETF
N33_HUMAN             AFCN------KLEFSMVDYDEG----TDVFQQLNMNSAETF
DROS._CG7830          TYSN------KLEFAMVDFDDG----SEVFQLLRLNTAEVF
Celegans_g304348      SEGDRR----KVEFGIVDYEDA----PQIFQQMNLNTAEIL
Yeast_Ost3p           PDAKSSNSDTSIEFTKVNLEDPSKTIPKADQFFQLNNVERL
Yeast_Ost6p           PQSLN-------LEFTVDVNEV----PQLVKDLKLQNVEHL

IAG2_HUMAN            INEPAK-GKPKRGDTYELQVRG--FSAEQIARWIADR----
N33_HUMAN             MHEPPK-GRPKRADTFDLQRIG--FAAEQLAKWIADR----
DROS._CG7830          MHEPAK-GKPKGADTMDIHRVG--FAADSIAKFVAER----
Celegans_g304348      YHEGPKLGAKKRPEQMDFQRQG--FDADAIGRFVADQ----
Yeast_Ost3p           FIFKPNSPSILDHSVISISTDTGSERMKQIIQAIKQF----
Yeast_Ost6p           VVYPPAESNKQSQFEWKTSPFYQYSLVPENAENTLQFGDFL

IAG2_HUMAN            -TDVNIRVIRPENYAGPLMLGLLLAVIGGLVYLRRSNMEF-
N33_HUMAN             -TDVHIRVFRPENYSGTIALALLVSLVGGLLYLRRNNLEF-
DROS._CG7830          -TDITIRIFRPENYSGTVAMITLVALVGSFLYIRRNNLEF-
Celegans_g304348      -TEVHVRVIRPENYTAPVVIALFVALLLGMLYMKRNSLDF-
Yeast_Ost3p           -SQVNDFSLHLEMDWTPIITSTIITFITVLLFKKQSKLMFS
Yeast_Ost6p           AKILNISITVEQAFNVQEFVYYFVACMVVFIFIKKVILPKV
                                ******TM 1*******cccccccccc
```

```
IAG2_HUMAN            -LFNKTGWAFAALCFVLAMTSGQMWNHIRGPPYAHKNPHTG
N33_HUMAN             -IYNKTGWAMVSLCIVFAMTSGQMWNHIRGPPYAHKNPHNG
DROS._CG7830          -LYNKNLWGAIAVFFCFAMISGQMWNHIRGPPLVHKS-QNG
Celegans_g304348      -LFNRTVWGFVCLAITFIFMSGQMWNHIRGPPFMITNPNTK
Yeast_Ost3p           IISSRIIWATLSTFFIICMISAYMFNQLRNTQLAGVGPKGE
Yeast_Ost6p           TNKWKLFSMILSLGILLPSITGYKFVEMNAIPFIARDAKN-
                      ccccc*******TM 2******

IAG2_HUMAN            HVNYIHGSSQACFVAETHIVLLFNGGVTLGMVLLCEAATSD
N33_HUMAN             QVSYIHGSSQACFVAESHIILVLNAAITMGMVLLNEAATSK
DROS._CG7830          GVAYIHGSSQGCLVVETYIVMFLNAMIVLGMILLIESGTPK
Celegans_g304348      EPSFIHGSTQFCLIAETYIVGLLYALIAIGFICVNEAADQS
Yeast_Ost3p           VMYFLPNEFQHCFAIETQVMVLIYGTLAALVVVLVKGIQFL
Yeast_Ost6p           RIMYFSGGSGWCFGIEIFSVSLMYIVMSALSVLLIYVPKIS
                                 ******TM 3*****cccccccc

IAG2_HUMAN            MDIGKR----------------KIMCVAGIGLVVLFFSWML
N33_HUMAN             GDVGKR----------------RIICLVGLGLVVFFFSFLL
DROS._CG7830          AHN-KN----------------RIMAMTGLVLLTVFFSFLL
Celegans_g304348      NSKDRKNAGKKLNPLSLLNIPTNTLAIAGLVCICVFFSFLL
Yeast_Ost3p           RSHLYP-------ETKKAYFIDAILASFCALFIYVFFAALT
Yeast_Ost6p           CVSEKMR---------------GLLSSFLACVLFYFFSYFI
                      cccccccccccccccccccccc*****TM 4*****
                                               TF (3)

IAG2_HUMAN            SIFRSKYHGYPYSFLMS----------------
N33_HUMAN             SIFRSKYHGYPYSDLDFE-(1)-----------
DROS._CG7830          SVFRSKAQGYPYISCSNRIDCSPVPVQVHPISFL
Celegans_g304348      SVFRSKYRGYPYSFLFA----------------
Yeast_Ost3p           TVFTIKSPAYPFPLLRLSAPFK-----------
Yeast_Ost6p           SCYLIKNPGYPIVF-------------------

                            FLIK (2)
```

Figure 9

C-termini of N33 splice forms

```
N33_67811_Translated_-_Longe      LVSLVGGLLYLRRNNLEFIYNKTGHAMVSLCIVFAMTSGQMWNHIRGPPY
N33_67891011_Translated_-_Lo      LVSLVGGLLYLRRNNLEFIYNKTGHAMVSLCIVFAMTSGQMWNHIRGPPY
N33_678911_Translated_-_Long      LVSLVGGLLYLRRNNLEFIYNKTGHAMVSLCIVFAMTSGQMWNHIRGPPY
N33_611_Translated_-_Longest      LVSLVGGLLYLRRNNLEFIYNKTGHAMVSLCIVFAMTSGQMWNHIRGPPY
N33_68+911_Translated_-_Long      LVSLVGGLLYLRRNNLEFIYNKTGHAMVSLCIVFAMTSGQMWNHIRGPPY
N33_68+11_Translated_-_Longe      LVSLVGGLLYLRRNNLEFIYNKTGHAMVSLCIVFAMTSGQMWNHIRGPPY
                                  **************************************************

N33_67811_Translated_-_Longe      AHKNPHNGQVSYIHGSSQAQFVAESHTILVINAAFIMGMVLLNEAATSKG
N33_67891011_Translated_-_Lo      AHKNPHNGQVSYIHGSSQAQFVAESHTILVINAAFIMGMVLLNEAATSKG
N33_678911_Translated_-_Long      AHKNPHNGQVSYIHGSSQAQFVAESHTILVINAAFIMGMVLLNEAATSKG
N33_611_Translated_-_Longest      AHKNPHNGQV---------------------------------------
N33_68+911_Translated_-_Long      AHKNPHNGQVFNHSG----TLCSEPEKLIDFIHIYVYG--FLDNYCRCRY
N33_68+11_Translated_-_Longe      AHKNPHNGQVFNHSG----TLCSEPEKLIDFIHIYVYG--FLDNYCRCRY
                                  **********

N33_67811_Translated_-_Longe      DVGKRR------------------------------------
N33_67891011_Translated_-_Lo      DVGKRRIICLVGLGLVVFFTSFLLSIFRSKYHGYPYS
N33_678911_Translated_-_Long      DVGKRRIICLVGLGLVVFFTSFLLSIFRSKYHGYPYS-
N33_611_Translated_-_Longest      ------------------------------------------
N33_68+911_Translated_-_Long      HHGDGSSK----------------------------------
N33_68+11_Translated_-_Longe      HHGDGSSK----------------------------------
```

Figure 10

Published GRIK4 nucleic acid sequence (accession NM_014619).

```
   1 atgccccgcg tctcggcgcc tttggtgctg cttcctgcgt ggctcgtgat ggtcgcctgc
  61 agcccgcact ccttgaggat cgctgctatc ttggacgacc ccatggagtg cagcagaggg
 121 gagcggctct ccatcaccct ggccaagaac cgcatcaacc gcgctcctga gaggctgggc
 181 aaggccaagg tcgaagtgga catctttgag cttctcagag acagcgagta cgagactgca
 241 gaaaccatgt gtcagatcct ccccaagggg gtggtcgctg tcctcggacc atcgtccagc
 301 ccagcctcca gctccatcat cagcaacatc tgtggagaga aggaggtccc tcacttcaaa
 361 gtggccccag aggagttcgt caagttccag ttccagagat tcacaaccct gaacctccac
 421 cccagcaaca ctgacatcag cgtggctgta gctgggatcc tgaacttctt caactgcacc
 481 accgcctgcc tcatctgtgc caaagcagaa tgccttttaa acctagagaa gctgctccgg
 541 caattcctta tctccaagga cacgctgtcc gtccgcatgc tggatgacac ccgggacccc
 601 accccgctcc tcaaggagat ccgggacgac aagaccgcca ccatcatcat ccacgccaac
 661 gcctccatgt cccacaccat cctcctgaag gcagccgaac ttgggatggt gtcagcctat
 721 tacacataca tcttcactaa tctggagttc tcactccaga gaacggacag ccttgtggat
 781 gatcgtgtca acatcctggg attttccatt ttcaaccaat cccatgcttt cttccaagag
 841 tttgcccaga gcctcaacca gtcctggcag gagaactgtg accatgtgcc cttcactggg
 901 cctgcgctct cctcggccct gctgtttgat gctgtctatg ctgtggtgac tgcggtgcag
 961 gaactgaacc ggagccaaga gatcggcgtg aagcccttgt cctgcggctc ggcccagatc
1021 tggcagcacg gcaccagcct catgaactac ctgcgcatgg tagaattgga aggtcttacc
1081 ggccacattg aattcaacag caaaggccag aggtccaact acgctttgaa aatcttacag
1141 ttcacaagga atggttttcg gcagatcggc cagtggcacg tggcagaggg cctcagcatg
1201 gacagccacc tctatgcctc caacatctcg gacactctct tcaacaccac cctggtcgtc
1261 accaccatcc tggaaaaccc atatttaatg ctgaagggga accaccagga gatggaaggc
1321 aatgaccgct acgagggctt ctgtgtggac atgctcaagg agctggcaga gatcctccga
1381 ttcaactaca agatccgcct ggttggggat ggcgtgtacg gcgttcccga ggccaacggc
1441 acctggacgg gaatggtcgg ggagctgatc gctaggaaag cagatctggc tgtggcaggc
1501 ctcaccatta cagctgaacg ggagaaggtg attgatttct ctaagccatt catgactctg
1561 ggaattagca ttctttaccg cattcatatg ggacgcaaac ccggctattt ctccttcctg
1621 gacccatttt ctccgggcgt ctggctcttc atgcttctag cctatctggc cgtcagctgt
1681 gtcctcttcc tggtggctcg gttgacgccc tacgagtggt acagcccaca cccatgtgcc
1741 cagggccggt gcaacctcct ggtgaaccag tactccctgg gcaacagcct ctggtttccg
1801 gtcggggggt tcatgcagca gggctccacc atcgcccctc gcgccttatc cacccgctgt
1861 gtcagtggcg tctggtgggc attcacgctg atcatcatct catcctacac ggccaacctg
1921 gcagccttcc tgaccgtgca gcgcatggat gtgcccattg agtcagtgga tgacctggct
1981 gaccagaccg ccattgaata tggcacaatt cacggaggct ccagcatgac cttcttccaa
2041 aattcccgct accagaccta ccaacgcatg tggaattaca tgtattccaa gcagcccagc
2101 gtgttcgtga agagcacaga ggagggaatc gccagggtgt tgaattccaa ctacgccttc
2161 ctcctggaat ccaccatgaa cgagtactat cggcagcgaa actgcaacct cactcagatt
2221 gggggcctgc tggacaccaa gggctatggg attggcatgc cagtcggctc ggtttccgg
2281 gacgagtttg atctggccat tctccagctg caggagaaca accgcctgga gatcctgaag
2341 cgcaaatggt gggaaggagg gaagtgcccc aaggaggaag atcacagagc taaaggcctg
2401 ggaatggaga atattggtgg aatctttgtg gttcttattt gtggcttaat cgtggccatt
2461 tttatggcta tgttggagtt tttatggact ctcagacact cagaagcaac tgaggtgtcc
2521 gtctgccagg agatggtgac cgagctgcgc agcattatcc tgtgtcagga cagtatccac
2581 ccccgccggc ggcgcgccgc agtcccgccg ccccggcccc ccatccccga ggagcgccga
2641 ccgcggggca cggcgacgct cagcaacggg aagctgtgcg ggcaggggga gcccgaccag
2701 ctcgcgcaga gactggcgca ggaggccgcc ctggtggccc gcggctgcac gcacatccgc
2761 gtctgccccg agtgccgccg cttccagggc ctgcgggcac ggccgtcgcc cgccccgcagc
2821 gaggagagcc tggagtggga gaaaaccacc aacagcagcg agcccgagta g
```

Figure 11

Published GRIK4 protein sequence (accession NP_055434).

MPRVSAPLVLLPAWLVMVACSPHSLRIAAILDDPMECSRGERLSITLAKNRINRAPERL
GKAKVEVDIFELLRDSEYETAETMCQILPKGVVAVLGPSSSPASSSIISNICGEKEVPH
FKVAPEEFVKFQFQRFTTLNLHPSNTDISVAVAGILNFFNCTTACLICAKAECLLNLEK
LLRQFLISKDTLSVRMLDDTRDPTPLLKEIRDDKTATIIIHANASMSHTILLKAAELGM
VSAYYTYIFTNLEFSLQRTDSLVDDRVNILGFSIFNQSHAFFQEFAQSLNQSWQENCDH
VPFTGPALSSALLFDAVYAVVTAVQELNRSQEIGVKPLSCGSAQIWQHGTSLMNYLRMV
ELEGLTGHIEFNSKGQRSNYALKILQFTRNGFRQIGQWHVAEGLSMDSHLYASNISDTL
FNTTLVVTTILENPYLMLKGNHQEMEGNDRYEGFCVDMLKELAEILRFNYKIRLVGDGV
YGVPEANGTWTGMVGELIARKADLAVAGLTITAEREKVIDFSKPFMTLGISILYRIHMG
RKPGYFSFLDPFSPGVWLFMLLAYLAVSCVLFLVARLTPYEWYSPHPCAQGRCNLLVNQ
YSLGNSLWFPVGGFMQQGSTIAPRALSTRCVSGVWWAFTLIIISSYTANLAAFLTVQRM
DVPIESVDDLADQTAIEYGTIHGGSSMTFFQNSRYQTYQRMWNYMYSKQPSVFVKSTEE
GIARVLNSNYAFLLESTMNEYYRQRNCNLTQIGGLLDTKGYGIGMPVGSVFRDEFDLAI
LQLQENNRLEILKRKWWEGGKCPKEEDHRAKGLGMENIGGIFVVLICGLIVAIFMAMLE
FLWTLRHSEATEVSVCQEMVTELRSIILCQDSIHPRRRRAAVPPPRPPIPEERRPRGTA
TLSNGKLCGAGEPDQLAQRLAQEAALVARGCTHIRVCPECRRFQGLRARPSPARSEESL
EWEKTTNSSEPE

Figure 12

Cytogenetic Position	Description	Breakpoint YAC Clones	Breakpoint BAC Clones (Acc. No.)
2p12	Inversion breakpoint	915_f_7	-
2q32.1	Inversion breakpoint	941_h_12	RP11-358M9 (AC020595)
2q21.3	Translocation breakpoint	766_c_12	RP11-250H22 (AC011996)
11q23.3	Upper insertion breakpoint	936_d_9	RP11-89P5 (AC009641)
11q24.2	Translocation/Insertion breakpoint	749_d_2	RP11-687M24 (AP001007)

Figure 13

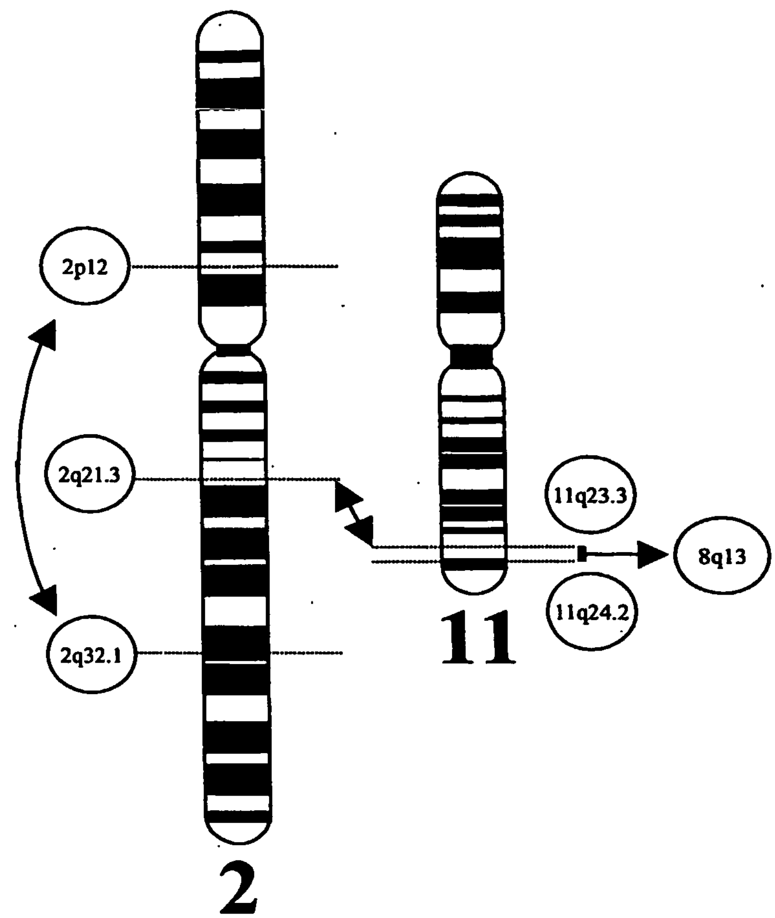


Figure 14

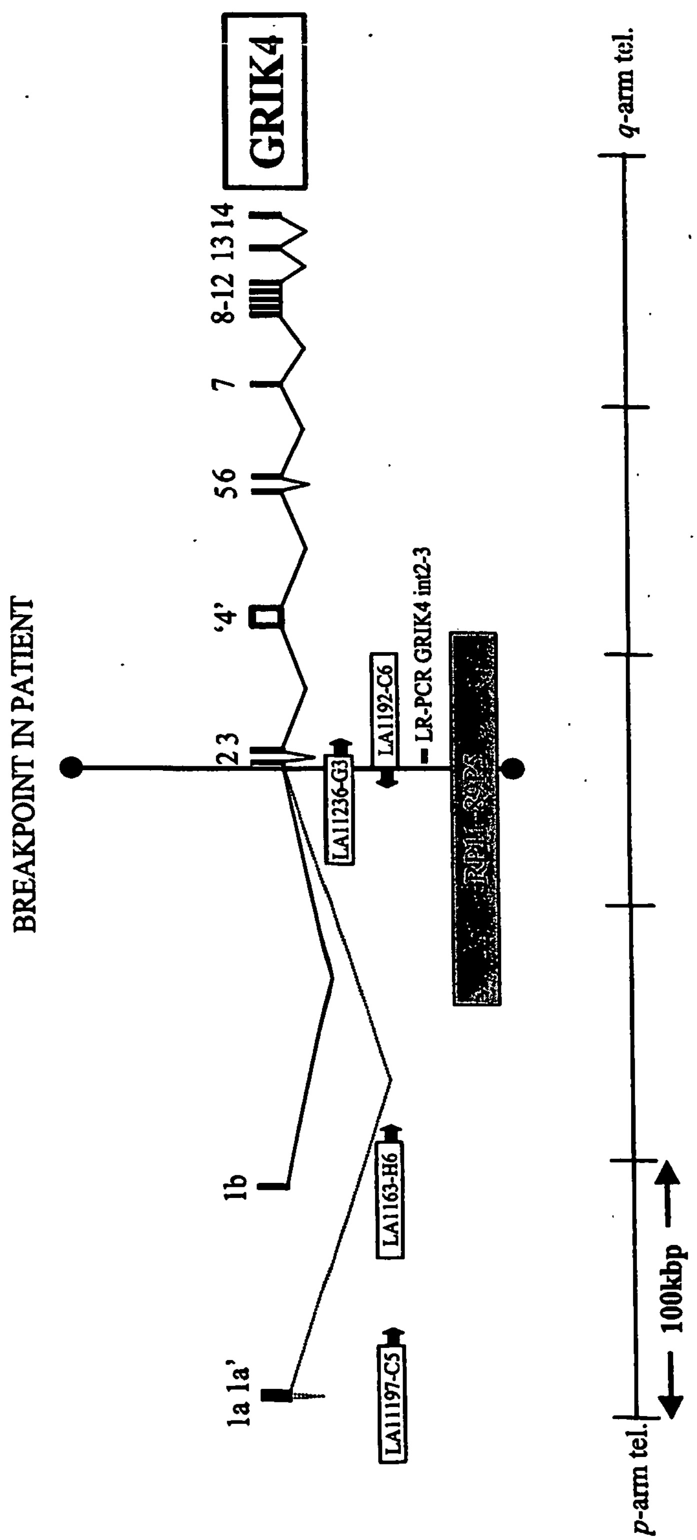

GRIK4
BREAKPOINT IN PATIENT
1a 1a'
1b
2 3
'4'
56
7
8-12 13 14
LA11197-C5
LA11163-H6
LA11236-G3
LA11192-C6
LR-PCR GRIK4 int2-3
p-arm tel.
q-arm tel.
100kbp

Figure 15

Exon 1a

```
GCGTGGTAGCATGTGCCTGTAATCCCAGTGCTTTGGGACACCGAGGCAGGAGGATCACT
CGAGCCCAGGAGTGCGAGGCTGCAgtgagttatgatcatac
```

Exon 1a'

```
agatttgtcttctctgccagGTGACGCTAGACTTCAGGAAGACCCCCCATTTCTGCTCC
ACTCCTGGGCTTGGAGAAGAGTACAGCTGCTCTTGACTGGTGGGACCTTTTGCTGGCTA
GGGGTGATGGGAGAAGCAAGAGAGGGATCCACACACCTGCGCTTAGCTTTCTATGACCT
GGGCGGATGGAGGCCAAAGgtaaggtgggatgaga
      M  E  A  K  A
```

Exon 1b

```
CCATGAGGATTCATAGAAGATGCCCCGCGTCTCGGCGCCTTTGGTGCTGCTTCCTGCGT
                  M  P  R  V  S  A  P  L  V  L  L  P  A  W
GGCTCGTGATGGTCGCCTGCAGCCCGCACTCCTTGAGGATCGgtaagtgtggcccagct
 L  V  M  V  A  C  S  P  H  S  L  R  I  A
```

Exon 2

```
gaaaccccccccagCTGCTATCTTGGACGACCCCATGGAGTGCAGCAGAGGGGAGCGGC
                A  I  L  D  D  P  M  E  C  S  R  G  E  R  L
TCTCCATCACCCTGGCCAAGAACCGCATCAACCGCGCTCCTGAGAGGCTGGGCAAGGCC
  S  I  T  L  A  K  N  R  I  N  R  A  P  E  R  L  G  K  A
AAGGTCGAAGTGGACATCTTTGAGCTTCTCAGAGACAGCGAGTACGAGACTGCAGAAAC
K  V  E  V  D  I  F  E  L  L  R  D  S  E  Y  E  T  A  E  T

CAgtacgtagactggg
 M
```

Figure 16

Alternative nucleic acid sequence. Exons 1a-1a'-2-etc.

```
   1 gcgtggtagc atgtgcctgt aatcccagtg ctttgggaca ccgaggcagg aggatcactc
  61 gagcccagga gtgcgaggct gcagtgacgc tagacttcag gaagaccccc catttctgct
 121 ccactcctgg gcttggagaa gagtacagct gctcttgact ggtgggacct tttgctggct
 181 aggggtgatg ggagaagcaa gagagggatc cacacacctg cgcttagctt tctatgacct
 241 gggcggatgg aggccaaagc tgctatcttg gacgacccca tggagtgcag cagaggggag
 301 cggctctcca tcaccctggc caagaaccgc atcaaccgcg ctcctgagag gctgggcaag
 361 gccaaggtcg aagtggacat ctttgagctt ctcagagaca gcgagtacga gactgcagaa
 421 accatgtgtc agatcctccc caagggggtg gtcgctgtcc tcggaccatc gtccagccca
 481 gcctccagct ccatcatcag caacatctgt ggagagaagg aggtccctca cttcaaagtg
 541 gccccagagg agttcgtcaa gttccagttc cagagattca caaccctgaa cctccacccc
 601 agcaacactg acatcagcgt ggctgtagct gggatcctga acttcttcaa ctgcaccacc
 661 gcctgcctca tctgtgccaa agcagaatgc cttttaaacc tagagaagct gctccggcaa
 721 ttccttatct ccaaggacac gctgtccgtc cgcatgctgg atgacacccg ggaccccacc
 781 ccgctcctca aggagatccg ggacgacaag accgccacca tcatcatcca cgccaacgcc
 841 tccatgtccc acaccatcct cctgaaggca gccgaacttg ggatggtgtc agcctattac
 901 acatacatct tcactaatct ggagttctca ctccagagaa cggacagcct tgtggatgat
 961 cgtgtcaaca tcctgggatt ttccattttc aaccaatccc atgctttctt ccaagagttt
1021 gcccagagcc tcaaccagtc ctggcaggag aactgtgacc atgtgccctt cactgggcct
1081 gcgctctcct cggccctgct gtttgatgct gtctatgctg tggtgactgc ggtgcaggaa
1141 ctgaaccgga gccaagagat cggcgtgaag cccttgtcct gcggctcggc ccagatctgg
1201 cagcacggca ccagcctcat gaactacctg cgcatggtag aattggaagg tcttaccggc
1261 cacattgaat tcaacagcaa aggccagagg tccaactacg ctttgaaaat cttacagttc
1321 acaaggaatg gttttcggca gatcggccag tggcacgtgg cagagggcct cagcatggac
1381 agccacctct atgcctccaa catctcggac actctcttca acaccaccct ggtcgtcacc
1441 accatcctgg aaaacccata tttaatgctg aaggggaacc accaggagat ggaaggcaat
1501 gaccgctacg agggcttctg tgtggacatg ctcaaggagc tggcagagat cctccgattc
1561 aactacaaga tccgcctggt tggggatggc gtgtacggcg ttcccgaggc caacggcacc
1621 tggacgggaa tggtcgggga gctgatcgct aggaaagcag atctggctgt ggcaggcctc
1681 accattacag ctgaacggga gaaggtgatt gatttctcta agccattcat gactctggga
1741 attagcattc tttaccgcat tcatatggga cgcaaacccg gctatttctc cttcctggac
1801 ccattttctc cgggcgtctg gctcttcatg cttctagcct atctggccgt cagctgtgtc
1861 ctcttcctgg tggctcggtt gacgccctac gagtggtaca gcccacaccc atgtgcccag
1921 ggccggtgca acctcctggt gaaccagtac tccctgggca acagcctctg gtttccggtc
1981 ggggggttca tgcagcaggg ctccaccatc gcccctcgcg ccttatccac ccgctgtgtc
2041 agtggcgtct ggtgggcatt cacgctgatc atcatctcat cctacacggc caacctggca
2101 gccttcctga ccgtgcagcg catggatgtg cccattgagt cagtggatga cctggctgac
2161 cagaccgcca ttgaatatgg cacaattcac ggaggctcca gcatgacctt cttccaaaat
2221 tcccgctacc agacctacca acgcatgtgg aattacatgt attccaagca gcccagcgtg
2281 ttcgtgaaga gcacagagga gggaatcgcc agggtgttga attccaacta cgccttcctc
2341 ctggaatcca ccatgaacga gtactatcgg cagcgaaact gcaacctcac tcagattggg
2401 ggcctgctgg acaccaaggg ctatgggatt ggcatgccag tcggctcggt tttccgggac
2461 gagtttgatc tggccattct ccagctgcag gagaacaacc gcctggagat cctgaagcgc
2521 aaatggtggg aaggagggaa gtgccccaag gaggaagatc acagagctaa aggcctggga
2581 atggagaata ttggtggaat ctttgtggtt cttatttgtg gcttaatcgt ggccattttt
2641 atggctatgt tggagttttt atggactctc agacactcag aagcaactga ggtgtccgtc
2701 tgccaggaga tggtgaccga gctgcgcagc attatcctgt gtcaggacag tatccacccc
2761 cgccggcggc gcgccgcagt cccgccgccc cggcccccca tccccgagga gcgccgaccg
2821 cggggcacgg cgacgctcag caacgggaag ctgtgcgggg caggggagcc cgaccagctc
2881 gcgcagagac tggcgcagga ggccgccctg gtggcccgcg gctgcacgca catccgcgtc
2941 tgccccgagt gccgccgctt ccagggcctg cgggcacggc cgtcgcccgc ccgcagcgag
3001 gagagcctgg agtgggagaa aaccaccaac agcagcgagc ccgagtag
```

Figure 17

Complete alternative protein sequence

MEAKAAILDDPMECSRGERLSITLAKNRINRAPERLGKAKVEVDIFELLRDSEYETAET
MCQILPKGVVAVLGPSSSPASSSIISNICGEKEVPHFKVAPEEFVKFQFQRFTTLNLHP
SNTDISVAVAGILNFFNCTTACLICAKAECLLNLEKLLRQFLISKDTLSVRMLDDTRDP
TPLLKEIRDDKTATIIIHANASMSHTILLKAAELGMVSAYYTYIFTNLEFSLQRTDSLV
DDRVNILGFSIFNQSHAFFQEFAQSLNQSWQENCDHVPFTGPALSSALLFDAVYAVVTA
VQELNRSQEIGVKPLSCGSAQIWQHGTSLMNYLRMVELEGLTGHIEFNSKGQRSNYALK
ILQFTRNGFRQIGQWHVAEGLSMDSHLYASNISDTLFNTTLVVTTILENPYLMLKGNHQ
EMEGNDRYEGFCVDMLKELAEILRFNYKIRLVGDGVYGVPEANGTWTGMVGELIARKAD
LAVAGLTITAEREKVIDFSKPFMTLGISILYRIHMGRKPGYFSFLDPFSPGVWLFMLLA
YLAVSCVLFLVARLTPYEWYSPHPCAQGRCNLLVNQYSLGNSLWFPVGGFMQQGSTIAP
RALSTRCVSGVWWAFTLIIISSYTANLAAFLTVQRMDVPIESVDDLADQTAIEYGTIHG
GSSMTFFQNSRYQTYQRMWNYMYSKQPSVFVKSTEEGIARVLNSNYAFLLESTMNEYYR
QRNCNLTQIGGLLDTKGYGIGMPVGSVFRDEFDLAILQLQENNRLEILKRKWWEGGKCP
KEEDHRAKGLGMENIGGIFVVLICGLIVAIFMAMLEFLWTLRHSEATEVSVCQEMVTEL
RSIILCQDSIHPRRRRAAVPPPRPPIPEERRPRGTATLSNGKLCGAGEPDQLAQRLAQE
AALVARGCTHIRVCPECRRFQGLRARPSPARSEESLEWEKTTNSSEPE

Figure 18

NPAS3 (NM_022123) nucleic acid sequence (spliceform 1b-3-4etc)

```
   1 ccacgcgtcc gacgcccccc acccgggagg ggggagagag gcaaaaagta agagaggaaa
  61 aaaaatagca ggaagatggc gcccaccaag cccagctttc agcaggatcc ttccaggcga
 121 gaacgtttac aagcattgag aaaggagaaa tcccgagatg ctgctcgctc ccgccgggga
 181 aaagaaaact ttgagttcta tgaattggcc aagttgttgc ctcttcctgc agccattacc
 241 agccagctcg acaaggcatc catcattcga cttacaatta gctatctgaa aatgagggac
 301 tttgctaacc aggggaccc tccgtggaac ttgcgaatgg aaggccctcc acctaacaca
 361 tcagtaaaag gtgcacagcg aaggagaagc cccagtgcac tagccattga agtatttgaa
 421 gcacatttgg gaagccacat tttgcagtcc ctggatggct ttgtatttgc actaaatcag
 481 gaaggaaaat ttttgtacat ttccgaaaca gtctccatct acctaggcct ctcacaagtg
 541 gagctgacag gcagcagtgt ctttgactat gtccaccccg gagatcacgt ggagatggct
 601 gagcagctgg gcatgaagct ccccccctggg cggggtctcc tgtcacaggg cactgctgag
 661 gacggagcca gctcagcatc ttcctcctct cagtcggaga cccccgagcc agtggagtca
 721 accagcccca gtctgctaac cactgacaac actcttgagc gttccttttt catccgaatg
 781 aaatctactc tgaccaaacg cggtgtgcac atcaaatcat caggatataa ggtgattcac
 841 ataacaggcc ggctacgcct gagagtgtcg ctgtcccacg ggaggaccgt ccccagccaa
 901 atcatgggtc tcgtggttgt tgcgcatgcc ttgcctcccc ctacgatcaa tgaagtcaga
 961 attgactgcc atatgttcgt cactcgagta aatatggacc tcaatatcat ttactgtgaa
1021 aataggatta gtgattatat ggatctgacc cctgtagata tcgtagggaa gagatgctac
1081 cacttcatcc atgctgaaga cgtggagggc atcaggcaca gtcacttgga cttgctgaat
1141 aagggtcagt gtgtgacaaa gtactatcgc tggatgcaga agaacggagg atatatttgg
1201 atacagtcca gtgccaccat agctattaat gccaagaatg caaatgaaaa gaatatcatc
1261 tgggtgaatt accttcttag caatcctgag tacaaggaca cacccatgga catcgcacag
1321 ctcccccatc tgccggagaa aacttccgaa tcctcggaga catccgactc tgagtcagac
1381 tctaaagaca cctcaggtat tacagaggac aacgagaact ccaagtccga cgagaagggg
1441 aaccagtccg agaacagcga agacccggag cccgaccgga agaagtcggg caacgcgtgt
1501 gacaacgaca tgaactgcaa cgacgacggc cacagctcca gtaacccgga cagccgcgac
1561 agcgacgaca gcttcgagca ctcggacttt gagaacccca aggcgggcga ggacggcttc
1621 ggtgctctgg gcgcgatgca gatcaaggtg gagcgctacg tggagagcga gtcggacctg
1681 cggctgcaga actgcgagtc actcacgtcc gacagcgcca aggactcgga cagcgcaggc
1741 gaggcgggcg cgcaggcctc cagcaagcac cagaagcgca agaaaaggcg gaaacggcaa
1801 aagggcggca gcgccagccg ccggcgcctg tccagcgcgt cgagcccagg cggcctggac
1861 gcgggcctgg tggagccccc gcggctgctg tcctccccca acagtgcctc ggtgctcaag
1921 atcaagacgg agatctcaga acccatcaat ttcgacaatg acagcagcat ctggaactac
1981 ccgcccaacc gggagatctc caggaacgag tcccccctaca gcatgaccaa gcccccccagc
2041 tctgagcact tcccgtcccc gcagggcggc ggcggtgggg gtggcggtgg cgggggggctg
2101 cacgtggcca ttcccgactc ggtcctcacc ccgcccggcg ccgacggcgc ggccgcccgc
2161 aagactcagt tcggcgcctc ggccaccgcg gccctggccc ccgtcgcctc cgacccgctg
2221 tcacccccgc tctcggcgtc cccgcgggac aagcaccccg ggaacggcgg cgggggcggg
2281 ggcgggggcg gcggcgcggg gggcggcggc cccagcgcgt ccaactcctt gctgtacact
2341 ggggacctgg aggcgctgca gaggttgcag gcgggcaacg tcgtgctccc gctggtgcac
2401 agggtgaccg ggaccctggc cgccaccagc acggccgcgc agagggtcta caccacgggc
2461 accatccgct acgcgcccgc cgaggtgacc ctggccatgc agagcaacct gctgcccaac
2521 gcgcacgctg ttaacttcgt ggacgttaac agccccggct ttggcctcga ccccaagacg
2581 cccatggaga tgctctacca ccacgtgcac cggctcaaca tgtcaggacc gttcggcggc
2641 gcagtgagcg cagctagcct gacgcagatg cccgccggca acgtgttcac cacggccgag
2701 ggactcttct ccacgctgcc cttccccgtc tacagcaacg gcatccacgc ggcacagact
2761 ctggagcgca aggaggactg aggcgccgcc cgtcctgggc ccggccaggc cccgcttgga
2821 ggaggcatcg tcggcatttt cgtttagacc tttaattcta gcactttgaa ttcgagcagg
```

```
2881 tcagcgtctt ctctcgccac gacggtcccc attccacccc ctctttcttt cacctgactt
2941 attctttcgt gtaaagatat gtttattttt tgccttcaga gggtcagacg accagttgcc
3001 tgccgttttg tcttcttcta aggtgtgtgt tgggttgttt tgctttcctt tgcatcttta
3061 ttaagatgtc tttcatgtgt atatgcctct gccatagaat actcagtctt gtggtcaaga
3121 gagttctcaa gtgacaacca ttggggtttc ttcataaaga tcttgatatg atcaagatgg
3181 aaagagacaa gcataaacaa tgtgccctgt ttgactaagt caaatgaaat agggtggttt
3241 ttgtttctgt tcctaattcc tttaaaaaat agggggaata gtattttaga attttatgca
3301 gaatttaatt ctctttttac ggttaagatt ttaagatttt cttacttgca cataaaaata
3361 atttgggttc ttaaacttaa tttctggcct gtgactagaa tgtttaaaaa aaaaaaaaac
3421 cctcgtgc
```

Figure 19

NPAS3 protein sequence (spliceform 1b-3-4etc.)

```
MAPTKPSFQQDPSRRERLQALRKEKSRDAARSRRGKENFEFYELAKLLPLPAAITSQLD
KASIIRLTISYLKMRDFANQGDPPWNLRMEGPPPNTSVKGAQRRRSPSALAIEVFEAHL
GSHILQSLDGFVFALNQEGKFLYISETVSIYLGLSQVELTGSSVFDYVHPGDHVEMAEQ
LGMKLPPGRGLLSQGTAEDGASSASSSSQSETPEPVESTSPSLLTTDNTLERSFFIRMK
STLTKRGVHIKSSGYKVIHITGRLRLRVSLSHGRTVPSQIMGLVVVAHALPPPTINEVR
IDCHMFVTRVNMDLNIIYCENRISDYMDLTPVDIVGKRCYHFIHAEDVEGIRHSHLDLL
NKGQCVTKYYRWMQKNGGYIWIQSSATIAINAKNANEKNIIWVNYLLSNPEYKDTPMDI
AQLPHLPEKTSESSETSDSESDSKDTSGITEDNENSKSDEKGNQSENSEDPEPDRKKSG
NACDNDMNCNDDGHSSSNPDSRDSDDSFEHSDFENPKAGEDGFGALGAMQIKVERYVES
ESDLRLQNCESLTSDSAKDSDSAGEAGAQASSKHQKRKKRRKRQKGGSASRRRLSSASS
PGGLDAGLVEPPRLLSSPNSASVLKIKTEISEPINFDNDSSIWNYPPNREISRNESPYS
MTKPPSSEHFPSPQGGGGGGGGGGGLHVAIPDSVLTPPGADGAAARKTQFGASATAALA
PVASDPLSPPLSASPRDKHPGNGGGGGGGGGAGGGGPSASNSLLYTGDLEALQRLQAG
NVVLPLVHRVTGTLAATSTAAQRVYTTGTIRYAPAEVTLAMQSNLLPNAHAVNFVDVNS
PGFGLDPKTPMEMLYHHVHRLNMSGPFGGAVSAASLTQMPAGNVFTTAEGLFSTLPFPV
YSNGIHAAQTLERKED
```

Figure 20

NPAS3 nucleic acid sequence (spliceform incorporating exons 1a-2-3-4etc) similar to mouse cDNA with accession number NM_013780)

```
   1 ATGGGGAGGG CCGGCGCCGC GGCCAACGGC ACCCCGCAGA ACGTCCAGGG CATCACCTCC
  61 TACCAGCAGC GAATAACTGC CCAGCATCCT CTGCCCAACC AATCAGAATG TAGGAAAATC
 121 TACAGATATG ACGGAATCTA CTGTGAATCT ACCTACCAGA ATTTACAAGC ATTGAGAAAG
 181 GAGAAATCCC GAGATGCTGC TCGCTCCCGC CGGGGAAAAG AAAACTTTGA GTTCTATGAA
 241 TTGGCCAAGT TGTTGCCTCT TCCTGCAGCC ATTACCAGCC AGCTCGACAA GGCATCCATC
 301 ATTCGACTTA CAATTAGCTA TCTGAAAATG AGGGACTTTG CTAACCAGGG GGACCCTCCG
 361 TGGAACTTGC GAATGGAAGG CCCTCCACCT AACACATCAG TAAAAGGTGC ACAGCGAAGG
 421 AGAAGCCCCA GTGCACTAGC CATTGAAGTA TTTGAAGCAC ATTTGGGAAG CCACATTTTG
 481 CAGTCCCTGG ATGGCTTTGT ATTTGCACTA AATCAGGAAG GAAAATTTTT GTACATTTCC
 541 GAAACAGTCT CCATCTACCT AGGCCTCTCA CAAGTGGAGC TGACAGGCAG CAGTGTCTTT
 601 GACTATGTCC ACCCCGGAGA TCACGTGGAG ATGGCTGAGC AGCTGGGCAT GAAGCTCCCC
 661 CCTGGGCGGG GTCTCCTGTC ACAGGGCACT GCTGAGGACG GAGCCAGCTC AGCATCTTCC
 721 TCCTCTCAGT CGGAGACCCC CGAGCCAGTG GAGTCAACCA GCCCCAGTCT GCTAACCACT
 781 GACAACACTC TTGAGCGTTC CTTTTTCATC CGAATGAAAT CTACTCTGAC CAAACGCGGT
 841 GTGCACATCA AATCATCAGG ATATAAGGTG ATTCACATAA CAGGCCGGCT ACGCCTGAGA
 901 GTGTCGCTGT CCCACGGGAG GACCGTCCCC AGCCAAATCA TGGGTCTCGT GGTTGTTGCG
 961 CATGCCTTGC CTCCCCCTAC GATCAATGAA GTCAGAATTG ACTGCCATAT GTTCGTCACT
1021 CGAGTAAATA TGGACCTCAA TATCATTTAC TGTGAAAATA GGATTAGTGA TTATATGGAT
1081 CTGACCCCTG TAGATATCGT AGGGAAGAGA TGCTACCACT TCATCCATGC TGAAGACGTG
1141 GAGGGCATCA GGCACAGTCA CTTGGACTTG CTGAATAAGG GTCAGTGTGT GACAAAGTAC
1201 TATCGCTGGA TGCAGAAGAA CGGAGGATAT ATTTGGATAC AGTCCAGTGC CACCATAGCT
1261 ATTAATGCCA AGAATGCAAA TGAAAAGAAT ATCATCTGGG TGAATTACCT TCTTAGCAAT
1321 CCTGAGTACA AGGACACACC CATGGACATC GCACAGCTCC CCCATCTGCC GGAGAAAACT
1381 TCCGAATCCT CGGAGACATC CGACTCTGAG TCAGACTCTA AAGACACCTC AGGTATTACA
1441 GAGGACAACG AGAACTCCAA GTCCGACGAG AAGGGGAACC AGTCCGAGAA CAGCGAAGAC
1501 CCGGAGCCCG ACCGGAAGAA GTCGGGCAAC GCGTGTGACA ACGACATGAA CTGCAACGAC
1561 GACGGCCACA GCTCCAGTAA CCCGGACAGC CGCGACAGCG ACGACAGCTT CGAGCACTCG
1621 GACTTTGAGA ACCCCAAGGC GGGCGAGGAC GGCTTCGGTG CTCTGGGCGC GATGCAGATC
1681 AAGGTGGAGC GCTACGTGGA GAGCGAGTCG GACCTGCGGC TGCAGAACTG CGAGTCACTC
1741 ACGTCCGACA GCGCCAAGGA CTCGGACAGC GCAGGCGAGG CGGGCGCGCA GGCCTCCAGC
1801 AAGCACCAGA AGCGCAAGAA AAGGCGGAAA CGGCAAAAGG GCGGCAGCGC CAGCCGCCGG
1861 CGCCTGTCCA GCGCGTCGAG CCCAGGCGGC CTGGACGCGG GCCTGGTGGA GCCCCCGCGG
1921 CTGCTGTCCT CCCCCAACAG TGCCTCGGTG CTCAAGATCA AGACGGAGAT CTCAGAACCC
1981 ATCAATTTCG ACAATGACAG CAGCATCTGG AACTACCCGC CCAACCGGGA GATCTCCAGG
2041 AACGAGTCCC CCTACAGCAT GACCAAGCCC CCCAGCTCTG AGCACTTCCC GTCCCCGCAG
2101 GGCGGCGGCG GTGGGGTGG CGGTGGCGGG GGGCTGCACG TGGCCATTCC CGACTCGGTC
2161 CTCACCCCGC CCGGCGCCGA CGGCGCGGCC GCCCGCAAGA CTCAGTTCGG CGCCTCGGCC
2221 ACCGCGGCCC TGGCCCCCGT CGCCTCCGAC CCGCTGTCAC CCCCGCTCTC GGCGTCCCCG
2281 CGGGACAAGC ACCCCGGGAA CGGCGGCGGG GGCGGGGGCG GGGGCGGCGG CGCGGGGGGC
2341 GGCGGCCCCA GCGCGTCCAA CTCCTTGCTG TACACTGGGG ACCTGGAGGC GCTGCAGAGG
2401 TTGCAGGCGG GCAACGTCGT GCTCCCGCTG GTGCACAGGG TGACCGGGAC CCTGGCCGCC
2461 ACCAGCACGG CCGCGCAGAG GGTCTACACC ACGGGCACCA TCCGCTACGC GCCCGCCGAG
2521 GTGACCCTGG CCATGCAGAG CAACCTGCTG CCCAACGCGC ACGCTGTTAA CTTCGTGGAC
2581 GTTAACAGCC CCGGCTTTGG CCTCGACCCC AAGACGCCCA TGGAGATGCT CTACCACCAC
2641 GTGCACCGGC TCAACATGTC AGGACCGTTC GGCGGCGCAG TGAGCGCAGC TAGCCTGACG
2701 CAGATGCCCG CCGGCAACGT GTTCACCACG GCCGAGGGAC TCTTCTCCAC GCTGCCCTTC
2761 CCCGTCTACA GCAACGGCAT CCACGCGGCA CAGACTCTGG AGCGCAAGGA GGACTGAGGC
2821 GCCGCCCGTC CTGGGCCCGG CCAGGCCCCG CTTGGAGGAG GCATCGTCGG CATTTTCGTT
```

2881 TAGACCTTTA ATTCTAGCAC TTTGAATTCG AGCAGGTCAG CGTCTTCTCT CGCCACGACG
2941 GTCCCCATTC CACCCCCTCT T

Figure 21

NPAS3 protein sequence of spliceform incorporating exons 1a-2-3-4etc.

MGRAGAAANGTPQNVQGTHSYQQRLTAQHPLRNQSTQRKSVYRYDGLYQLSTYQNLQALR
KEKSRDAARSRRGKENFEFYELAKLLPLPAAITSQLDKASIIRLTISYLKMRDFANQGD
PPWNLRMEGPPPNTSVKGAQRRRSPSALAIEVFEAHLGSHILQSLDGFVFALNQEGKFL
YISETVSIYLGLSQVELTGSSVFDYVHPGDHVEMAEQLGMKLPPGRGLLSQGTAEDGAS
SASSSSQSETPEPVESTSPSLLTTDNTLERSFFIRMKSTLTKRGVHIKSSGYKVIHITG
RLRLRVSLSHGRTVPSQIMGLVVVAHALPPPTINEVRIDCHMFVTRVNMDLNIIYCENR
ISDYMDLTPVDIVGKRCYHFIHAEDVEGIRHSHLDLLNKGQCVTKYYRWMQKNGGYIWI
QSSATIAINAKNANEKNIIWVNYLLSNPEYKDTPMDIAQLPHLPEKTSESSETSDSESD
SKDTSGITEDNENSKSDEKGNQSENSEDPEPDRKKSGNACDNDMNCNDDGHSSSNPDSR
DSDDSFEHSDFENPKAGEDGFGALGAMQIKVERYVESESDLRLQNCESLTSDSAKDSDS
AGEAGAQASSKHQKRKKRRKRQKGGSASRRRLSSASSPGGLDAGLVEPPRLLSSPNSAS
VLKIKTEISEPINFDNDSSIWNYPPNREISRNESPYSMTKPPSSEHFPSPQGGGGGGGG
GGGLHVAIPDSVLTPPGADGAAARKTQFGASATAALAPVASDPLSPPLSASPRDKHPGN
GGGGGGGGGGAGGGGPSASNSLLYTGDLEALQRLQAGNVVLPLVHRVTGTLAATSTAAQ
RVYTTGTIRYAPAEVTLAMQSNLLPNAHAVNFVDVNSPGFGLDPKTPMEMLYHHVHRLN
MSGPFGGAVSAASLTQMPAGNVFTTAEGLFSTLPFPVYSNGIHAAQTLERKED

Figure 22

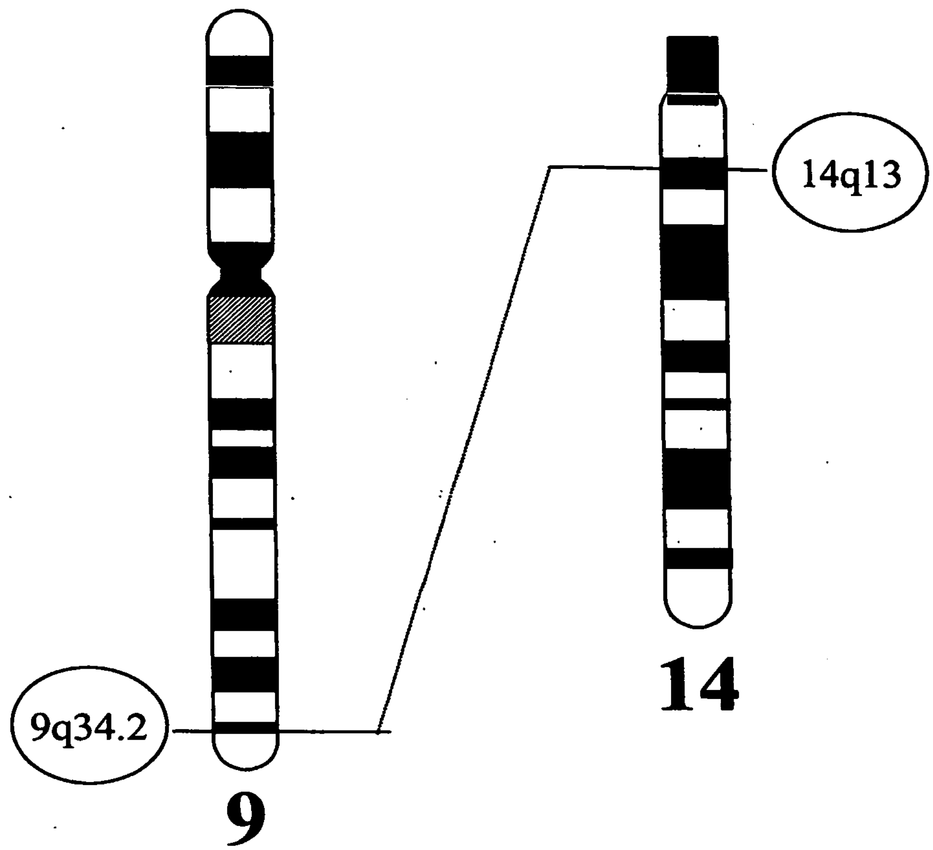


Figure 23

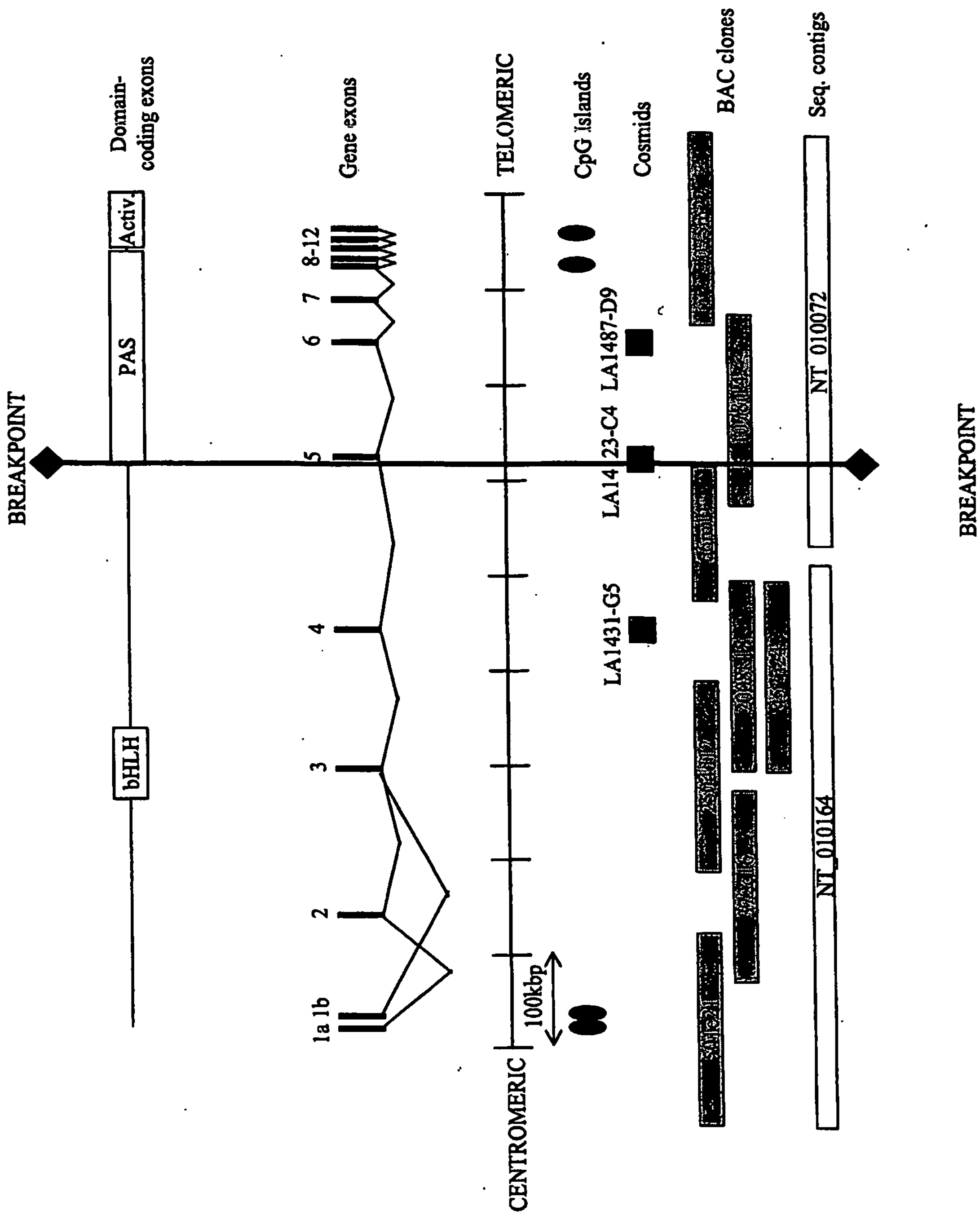

Domain-coding exons
Activ.
PAS
bHLH
Gene exons
8-12
7
6
5
4
3
2
1a 1b
TELOMERIC
CENTROMERIC
100kbp
CpG islands
Cosmids
LA1487-D9
LA1423-C4
LA1431-G5
BAC clones
Seq. contigs
NT 010072
NT_010164
BREAKPOINT
BREAKPOINT

Figure 24

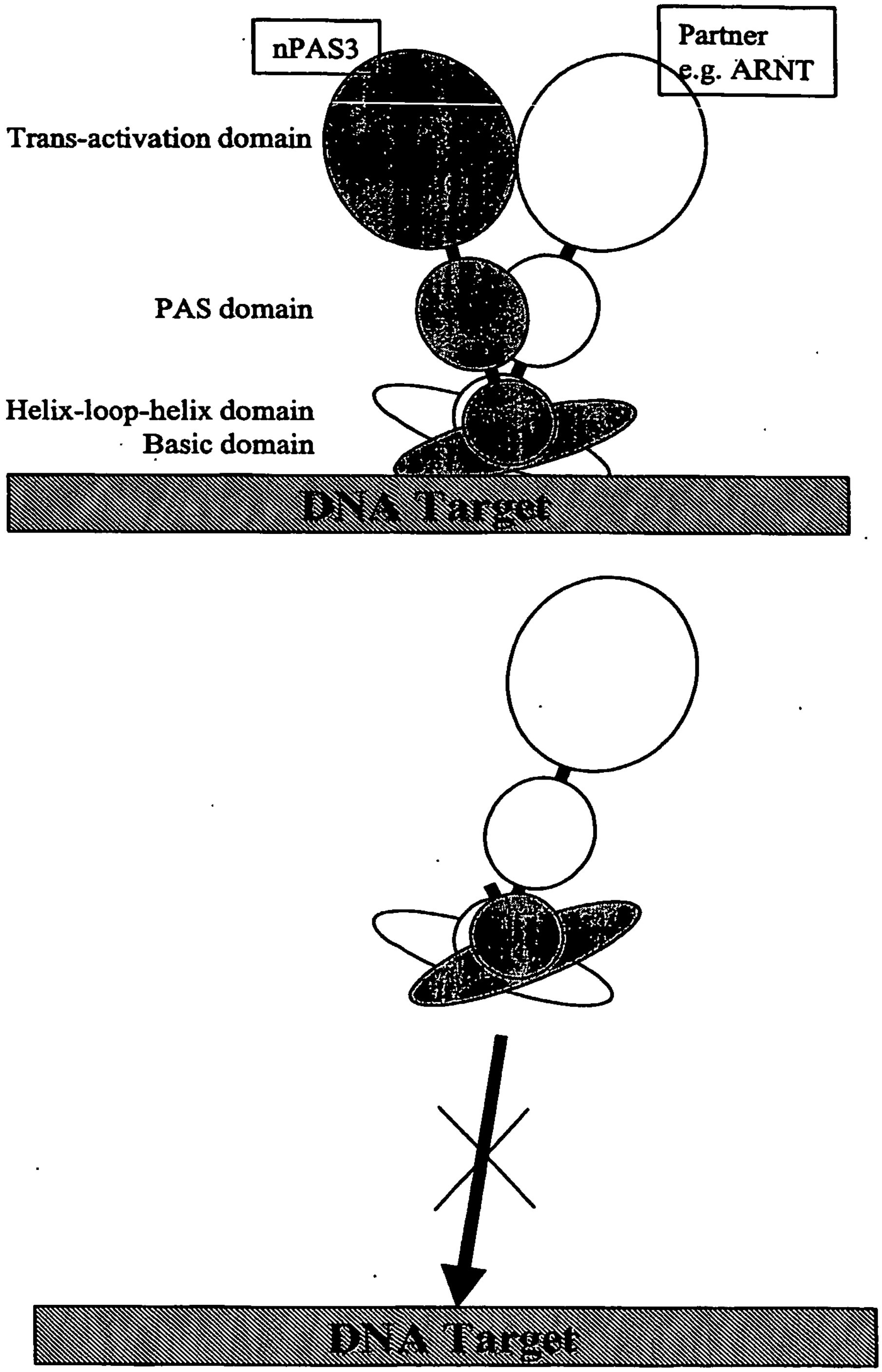

nPAS3
Partner
e.g. ARNT
Trans-activation domain
PAS domain
Helix-loop-helix domain
Basic domain
DNA Target
DNA Target

Figure 25
PDE4B1 (acc. L20966) Nucleic acid sequence

```
   1 gcggccgcgg cggtgcagca gaggcgcctc gggcaggagg agggcggctt ctgcgagggc
  61 agcctgaggt attaaaaagt gtcagcaaac tgcattgaat aacagacatc ctaagagggg
 121 atattttcca cctctataat gaagaaaagc aggagtgtga tgacggtgat ggctgatgat
 181 aatgttaaag attattttga atgtagcttg agtaaatcct acagttcttc cagtaacaca
 241 cttgggatcg acctctggag agggagaagg tgttgctcag gaaacttaca gttaccacca
 301 ctgtctcaaa gacagagtga aagggcaagg actcctgagg gagatggtat ttccaggccg
 361 accacactgc ctttgacaac gcttccaagc attgctatta caactgtaag ccaggagtgc
 421 tttgatgtgg aaaatggccc ttccccaggt cggagtccac tggatcccca ggccagctct
 481 tccgctgggc tggtacttca cgccaccttt cctgggcaca gccagcgcag agagtcattt
 541 ctctacagat cagacagcga ctatgacttg tcaccaaagg cgatgtcgag aaactcttct
 601 cttccaagcg agcaacacgg cgatgacttg attgtaactc cttttgccca ggtccttgcc
 661 agcttgcgaa gtgtgagaaa caacttcact atactgacaa accttcatgg tacatctaac
 721 aagaggtccc cagctgctag tcagcctcct gtctccagag tcaacccaca agaagaatct
 781 tatcaaaaat tagcaatgga aacgctggag gaattagact ggtgtttaga ccagctagag
 841 accatacaga cctaccggtc tgtcagtgag atggcttcta acaagttcaa aagaatgctg
 901 aaccgggagc tgacacacct ctcagagatg agccgatcag ggaaccaggt gtctgaatac
 961 atttcaaata ctttcttaga caagcagaat gatgtggaga tcccatctcc tacccagaaa
1021 gacagggaga aaaagaaaaa gcagcagctc atgacccaga taagtggagt gaagaaatta
1081 atgcatagtt caagcctaaa caatacaagc atctcacgct ttggagtcaa cactgaaaat
1141 gaagatcacc tggccaagga gctggaagac ctgaacaaat ggggtcttaa catctttaat
1201 gtggctggat attctcacaa tagaccccta acatgcatca tgtatgctat attccaggaa
1261 agagacctcc taaagacatt cagaatctca tctgacacat ttataaccta catgatgact
1321 ttagaagacc attaccattc tgacgtggca tatcacaaca gcctgcacgc tgctgatgta
1381 gcccagtcga cccatgttct cctttctaca ccagcattag acgctgtctt cacagatttg
1441 gagatcctgg ctgccatttt tgcagctgcc atccatgacg ttgatcatcc tggagtctcc
1501 aatcagtttc tcatcaacac aaattcagaa cttgctttga tgtataatga tgaatctgtg
1561 ttggaaaatc atcaccttgc tgtgggtttc aaactgctgc aagaagaaca ctgtgacatc
1621 ttcatgaatc tcaccaagaa gcagcgtcag acactcagga agatggttat tgacatggtg
1681 ttagcaactg atatgtctaa acatatgagc ctgctggcag acctgaagac aatggtagaa
1741 acgaagaaag ttacaagttc aggcgttctt ctcctagaca actataccga tcgcattcag
1801 gtccttcgca acatggtaca ctgtgcagac ctgagcaacc ccaccaagtc cttggaattg
1861 tatcggcaat ggacagaccg catcatggag gaatttttcc agcagggaga caaagagcgg
1921 gagaggggaa tggaaattag cccaatgtgt gataaacaca cagcttctgt ggaaaaatcc
1981 caggttggtt tcatcgacta cattgtccat ccattgtggg agacatgggc agatttggta
2041 cagcctgatg ctcaggacat tctcgatacc ttagaagata acaggaactg gtatcagagc
2101 atgatacctc aaagtccctc accaccactg gacgagcaga acagggactg ccagggtctg
2161 atggagaagt ttcagtttga actgactctc gatgaggaag attctgaagg acctgagaag
2221 gagggagagg gacacagcta tttcagcagc acaaagacgc tttgtgtgat tgatccagaa
2281 aacagagatt ccctgggaga gactgacata gacattgcaa cagaagacaa gtcccccgtg
2341 gatacataat ccccctctcc ctgtggagat gaacattcta tccttgatga gcatgccagc
2401 tatgtggtag ggccagccca ccatgggggc caagacctgc acaggacaag ggccacctgg
2461 cctttcagtt acttgagttt ggagtcagaa agcaagacca ggaagcaaat agcagctcag
2521 gaaatcccac ggttgacttg ccttgatggc aagcttggtg gagagggctg aagctgttgc
2581 tgggggccga ttctgatcaa gacacatggc ttgaaaatgg aagacacaaa actgagagat
2641 cattctgcac taagtttcgg gaacttatcc ccgacagtga ctgaactcac tgactaataa
2701 cttcatttat gaatcttctc acttgtccct ttgtctgcca acctgtgtgc ctttttttgta
2761 aaacattttc atgtctttaa aatgcctgtt gaatacctgg agtttagtat caacttctac
2821 acagataagc tttcaaagtt gacaaacttt tttgactctt tctggaaaag ggaaagaaaa
2881 tagtcttcct tctttcttgg gcaatatcct tcactttact acagttactt ttgcaaacag
2941 acagaaagga tacacttcta accacatttt acttccttcc cctgttgtcc agtccaactc
3001 cacagtcact cttaaaactt ctctctgttt gcctgcctcc aacagtactt ttaacttttt
```

```
3061 gctgtaaaca gaataaaatt gaacaaatta gggggtagaa aggagcagtg gtgtcgttca
3121 ccgtgagagt ctgcatagaa ctcagcagtg tgccctgctg tgtcttggac cctgcaatgc
3181 ggccgc
```

Figure 26

PDE4B1 Protein sequence

MKKSRSVMTVMADDNVKDYFECSLSKSYSSSSNTLGIDLWRGRRCCSGNLQLPPLSQRQ
SERARTPEGDGISRPTTLPLTTLPSIAITTVSQECFDVENGPSPGRSPLDPQASSSAGL
VLHATFPGHSQRRESFLYRSDSDYDLSPKAMSRNSSLPSEQHGDDLIVTPFAQVLASLR
SVRNNFTILTNLHGTSNKRSPAASQPPVSRVNPQEESYQKLAMETLEELDWCLDQLETI
QTYRSVSEMASNKFKRMLNRELTHLSEMSRSGNQVSEYISNTFLDKQNDVEIPSPTQKD
REKKKKQQLMTQISGVKKLMHSSSLNNTSISRFGVNTENEDHLAKELEDLNKWGLNIFN
VAGYSHNRPLTCIMYAIFQERDLLKTFRISSDTFITYMMTLEDHYHSDVAYHNSLHAAD
VAQSTHVLLSTPALDAVFTDLEILAAIFAAAIHDVDHPGVSNQFLINTNSELALMYNDE
SVLENHHLAVGFKLLQEEHCDIFMNLTKKQRQTLRKMVIDMVLATDMSKHMSLLADLKT
MVETKKVTSSGVLLLDNYTDRIQVLRNMVHCADLSNPTKSLELYRQWTDRIMEEFFQQG
DKERERGMEISPMCDKHTASVEKSQVGFIDYIVHPLWETWADLVQPDAQDILDTLEDNR
NWYQSMIPQSPSPPLDEQNRDCQGLMEKFQFELTLDEEDSEGPEKEGEGHSYFSSTKTL
CVIDPENRDSLGETDIDIATEDKSPVDT

Figure 27

PDE4B3 (acc. U85048) Nucleic acid sequence

```
   1 atgacagcaa aagattcttc aaaggaactt actgcttctg aacctgaggt ttgcataaag
  61 actttcaagg agcaaatgca tttagaactt gagcttccga gattaccagg aaacagacct
 121 acatctccta aaatttctcc acgcagttca ccaaggaact caccatgctt tttcagaaag
 181 ttactggtga ataaaagcat tcggcagcgt cgtcgcttca ctgtggctca tacatgcttt
 241 gatgtggaaa atggcccttc cccaggtcgg agtccactgg atccccaggc cagctcttcc
 301 gctgggctgg tacttcacgc cacctttcct gggcacagcc agcgcagaga gtcatttctc
 361 tacagatcag acagcgacta tgacttgtca ccaaaggcga tgtcgagaaa ctcttctctt
 421 ccaagcgagc aacacggcga tgacttgatt gtaactcctt ttgcccaggt ccttgccagc
 481 ttgcgaagtg tgagaaacaa cttcactata ctgacaaacc ttcatggtac atctaacaag
 541 aggtccccag ctgctagtca gcctcctgtc tccagagtca acccacaaga agaatcttat
 601 caaaaattag caatggaaac gctggaggaa ttagactggt gtttagacca gctagagacc
 661 atacagacct accggtctgt cagtgagatg gcttctaaca agttcaaaag aatgctgaac
 721 cgggagctga cacacctctc agagatgagc cgatcaggga accaggtgtc tgaatacatt
 781 tcaaatactt tcttagacaa gcagaatgat gtggagatcc catctcctac ccagaaagac
 841 agggagaaaa agaaaaagca gcagctcatg acccagataa gtggagtgaa gaaattaatg
 901 catagttcaa gcctaaacaa tacaagcatc tcacgctttg gagtcaacac tgaaaatgaa
 961 gatcacctgg ccaaggagct ggaagacctg aacaaatggg gtcttaacat ctttaatgtg
1021 gctggatatt ctcacaatag accccctaaca tgcatcatgt atgctatatt ccaggaaaga
1081 gacctcctaa agacattcag aatctcatct gacacattta taacctacat gatgacttta
1141 gaagaccatt accattctga cgtggcatat cacaacagcc tgcacgctgc tgatgtagcc
1201 cagtcgaccc atgttctcct ttctacacca gcattagacg ctgtcttcac agatttggag
1261 atcctggctg ccatttttgc agctgccatc catgacgttg atcatcctgg agtctccaat
1321 cagtttctca tcaacacaaa ttcagaactt gctttgatgt ataatgatga atctgtgttg
1381 gaaaatcatc accttgctgt gggtttcaaa ctgctgcaag aagaacactg tgacatcttc
1441 atgaatctca ccaagaagca gcgtcagaca ctcaggaaga tggttattga catggtgtta
1501 gcaactgata tgtctaaaca tatgagcctg ctggcagacc tgaagacaat ggtagaaacg
1561 aagaaagtta caagttcagg cgttcttctc ctagacaact ataccgatcg cattcaggtc
1621 cttcgcaaca tggtacactg tgcagacctg agcaacccca ccaagtcctt ggaattgtat
1681 cggcaatgga cagaccgcat catggaggaa tttttccagc agggagacaa agagcgggag
1741 aggggaatgg aaattagccc aatgtgtgat aaacacacag cttctgtgga aaaatcccag
1801 gttggtttca tcgactacat tgtccatcca ttgtgggaga catgggcaga tttggtacag
1861 cctgatgctc aggacattct cgatacctta gaagataaca ggaactggta tcagagcatg
1921 atacctcaaa gtccctcacc accactggac gagcagaaca gggactgcca gggtctgatg
1981 gagaagtttc agtttgaact gactctcgat gaggaagatt ctgaaggacc tgagaaggag
2041 ggagagggac acagctattt cagcagcaca aagacgcttt gtgtgattga tccagaaaac
2101 agagattccc tgggagagac tgacatagac attgcaacag aagacaagtc ccccgtggat
2161 aca
```

Figure 28

PDE4B3 Protein sequence

MEAKDGSKRLLASEPEVQLKTPKEQMHLEDEKPRLPGNKPTSPKTSPRSSPRNSPCLHK
KLLVNKSIPRQRRRFTVAHTCFDVENGPSPGRSPLDPQASSSAGLVLHATFPGHSQRRES
FLYRSDSDYDLSPKAMSRNSSLPSEQHGDDLIVTPFAQVLASLRSVRNNFTILTNLHGT
SNKRSPAASQPPVSRVNPQEESYQKLAMETLEELDWCLDQLETIQTYRSVSEMASNKFK
RMLNRELTHLSEMSRSGNQVSEYISNTFLDKQNDVEIPSPTQKDREKKKKQQLMTQISG
VKKLMHSSSLNNTSISRFGVNTENEDHLAKELEDLNKWGLNIFNVAGYSHNRPLTCIMY
AIFQERDLLKTFRISSDTFITYMMTLEDHYHSDVAYHNSLHAADVAQSTHVLLSTPALD
AVFTDLEILAAIFAAAIHDVDHPGVSNQFLINTNSELALMYNDESVLENHHLAVGFKLL
QEEHCDIFMNLTKKQRQTLRKMVIDMVLATDMSKHMSLLADLKTMVETKKVTSSGVLLL
DNYTDRIQVLRNMVHCADLSNPTKSLELYRQWTDRIMEEFFQQGDKERERGMEISPMCD
KHTASVEKSQVGFIDYIVHPLWETWADLVQPDAQDILDTLEDNRNWYQSMIPQSPSPPL
DEQNRDCQGLMEKFQFELTLDEEDSEGPEKEGEGHSYFSSTKTLCVIDPENRDSLGETD
IDIATEDKSPVDT

Figure 29

PDE4B2 (acc. NM_002600) Nucleic acid seqeunce

```
   1 gaattcctcc tctcttcacc ccgttagctg ttttcaatgt aatgctgccg tccttctctt
  61 gcactgcctt ctgcgctaac acctccattc ctgtttataa ccgtgtattt attacttaat
 121 gtatataatg taatgttttg taagttatta atttatatat ctaacattgc ctgccaatgg
 181 tggtgttaaa tttgtgtaga aaactctgcc taagagttac gacttttct tgtaatgttt
 241 tgtattgtgt attatataac ccaaacgtca cttagtagag acatatggcc cccttggcag
 301 agaggacagg ggtgggcttt tgttcaaagg gtctgccctt tccctgcctg agttgctact
 361 tctgcacaac ccctttatga accagtttc acccgaattt tgactgtttc atttagaaga
 421 aaagcaaaat gagaaaaagc tttcctcatt tctccttgag atggcaaagc actcagaaat
 481 gacatcacat accctaaaga accctgggat gactaaggca gagagagtct gagaaaactc
 541 tttggtgctt ctgcctttag ttttaggaca catttatgca gatgagctta taagagaccg
 601 ttccctccgc cttcttcctc agaggaagtt tcttggtaga tcaccgacac ctcatccagg
 661 cggggggttg gggggaaact tggcaccagc catcccaggc agagcaccac tgtgatttgt
 721 tctcctggtg gagagagctg gaaggaagga gccagcgtgc aaataatgaa ggagcacggg
 781 ggcaccttca gtagcaccgg aatcagcggt ggtagcggtg actctgctat ggacagcctg
 841 cagccgctcc agcctaacta catgcctgtg tgtttgtttg cagaagaatc ttatcaaaaa
 901 ttagcaatgg aaacgctgga ggaattagac tggtgtttag accagctaga gaccatacag
 961 acctaccggt ctgtcagtga gatggcttct aacaagttca aaagaatgct gaaccgggag
1021 ctgacacacc tctcagagat gagccgatca gggaaccagg tgtctgaata catttcaaat
1081 actttcttag acaagcagaa tgatgtggag atcccatctc ctacccagaa agacagggag
1141 aaaaagaaaa agcagcagct catgacccag ataagtggag tgaagaaatt aatgcatagt
1201 tcaagcctaa acaatacaag catctcacgc tttggagtca acactgaaaa tgaagatcac
1261 ctggccaagg agctggaaga cctgaacaaa tggggtctta acatctttaa tgtggctgga
1321 tattctcaca atagaccccct aacatgcatc atgtatgcta tattccagga aagagacctc
1381 ctaaagacat tcagaatctc atctgacaca tttataacct acatgatgac tttagaagac
1441 cattaccatt ctgacgtggc atatcacaac agcctgcacg ctgctgatgt agcccagtcg
1501 acccatgttc tcctttctac accagcatta gacgctgtct tcacagattt ggagatcctg
1561 gctgccattt ttgcagctgc catccatgac gttgatcatc ctggagtctc caatcagttt
1621 ctcatcaaca caaattcaga acttgctttg atgtataatg atgaatctgt gttggaaaat
1681 catcaccttg ctgtgggttt caaactgctg caagaagaac actgtgacat cttcatgaat
1741 ctcaccaaga agcagcgtca gacactcagg aagatggtta ttgacatggt gttagcaact
1801 gatatgtcta aacatatgag cctgctggca gacctgaaga caatggtaga aacgaagaaa
1861 gttacaagtt caggcgttct tctcctagac aactataccg atcgcattca ggtccttcgc
1921 aacatggtac actgtgcaga cctgagcaac cccaccaagt ccttggaatt gtatcggcaa
1981 tggacagacc gcatcatgga ggaatttttc cagcagggag acaaagagcg ggagagggga
2041 atggaaatta gcccaatgtg tgataaacac acagcttctg tggaaaaatc ccaggttggt
2101 ttcatcgact acattgtcca tccattgtgg gagacatggg cagatttggt acagcctgat
2161 gctcaggaca ttctcgatac cttagaagat aacaggaact ggtatcagag catgatacct
2221 caaagtccct caccaccact ggacgagcag aacagggact gccagggtct gatggagaag
2281 tttcagtttg aactgactct cgatgaggaa gattctgaag gacctgagaa ggagggagag
2341 ggacacagct atttcagcag cacaaagacg ctttgtgtga ttgatccaga aaacagagat
2401 tccctgggag agactgacat agacattgca acagaagaca agtcccccgt ggatacataa
2461 tccccctctc cctgtggaga tgaacattct atccttgatg agcatgccag ctatgtggta
2521 gggccagccc accatggggg ccaagacctg cacaggacaa gggccacctg gcctttcagt
2581 tacttgagtt tggagtcaga aagcaagacc aggaagcaaa tagcagctca ggaaatccca
2641 cggttgactt gccttgatgg caagcttggt ggagagggct gaagctgttg ctgggggccg
2701 attctgatca agacacatgg cttgaaaatg gaagacacaa aactgagaga tcattctgca
2761 ctaagtttcg ggaacttatc cccgacagtg actgaactca ctgactaata acttcattta
2821 tgaatcttct cacttgtccc tttgtctgcc aacctgtgtg ccttttttgt aaaacatttt
2881 catgtcttta aaatgcctgt tgaatacctg gagtttagta tcaacttcta cacagataag
2941 ctttcaaagt tgacaaactt ttttgactct ttctggaaaa gggaaagaaa atagtcttcc
```

3001 ttctttcttg ggcaatatcc ttcactttac tacagttact tttgcaaaca gacagaaagg
3061 atacacttct aaccacattt tacttccttc ccctgttgtc cagtccaact ccacagtcac
3121 tcttaaaact tctctctgtt tgcctgcctc caacagtact tttaactttt tgctgtaaac
3181 agaataaaat tgaacaaatt agggggtaga aaggagcagt ggtgtcgttc accgtgagag
3241 tctgcataga actcagcagt gtgccctgct gtgtcttgga ccctgccccc cacaggagtt
3301 gctacagtcc ctggccctgc ttcccatcct cctctcttca ccccgttagc tgttttcaat
3361 gtaatgctgc cgtccttctc ttgcactgcc ttctgcgcta acacctccat tcctgtttat
3421 aaccgtgtat ttattactta atgtatataa tgtaatgttt tgtaagttat taatttatat
3481 atctaacatt gcctgccaat ggtggtgtta aatttgtgta gaaaactctg cctaagagtt
3541 acgacttttt cttgtaatgt tttgtattgt gtattatata acccaaacgt cacttagtag
3601 agacatatgg ccccccttggc agagaggaca ggggtgggct tttgttcaaa gggtctgccc
3661 tttccctgcc tgagttgcta cttctgcaca accccttat gaaccagttt tggaaacaat
3721 attctcacat tagatactaa atggtttata ctgagtcttt tacttttgta tagcttgata
3781 ggggcagggg caatgggatg tagtttttac ccaggttcta tccaaatcta tgtgggcatg
3841 agttgggtta taactggatc ctactatcat tgtggctttg gttcaaaagg aaacactaca
3901 tttgctcaca gatgattctt ctgattcttc tgaatgctcc cgaactactg actttgaaga
3961 ggtagcctcc tgcctgccat taagcaggaa tgtcatgttc cagttcatta caaaagaaaa
4021 caataaaaca atgtgaattt ttataataaa aaaaaaaaaa aggaattc

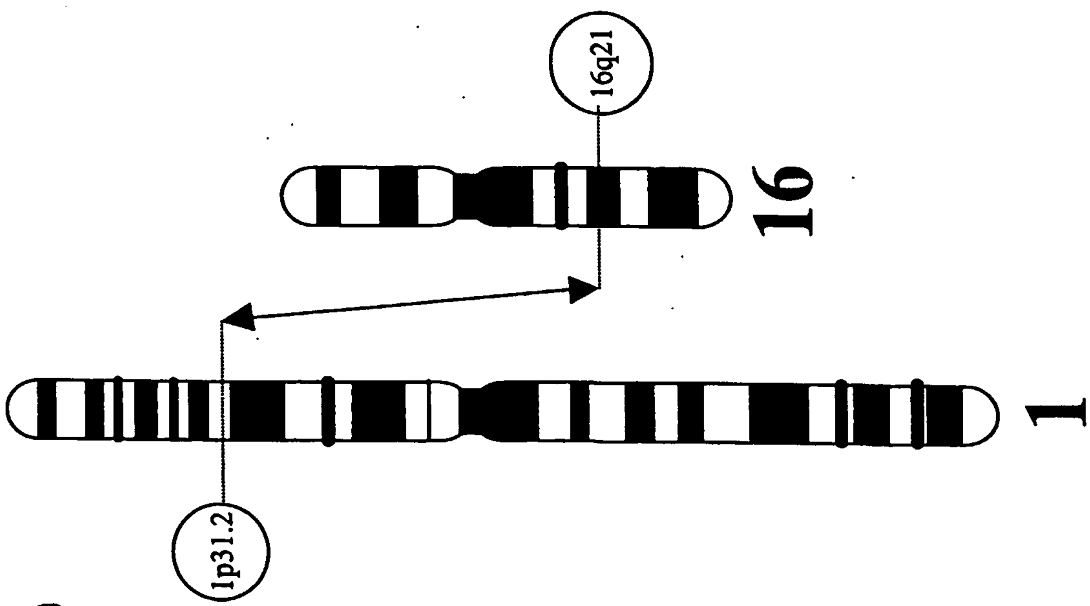
Figure 30

PDE4B2 Protein sequence

MKEHGGTFSSTGISGGSGDSAMDSLQPLQPNYMPVCLFAEESYQKLAMETLEELDWCLD
QLETIQTYRSVSEMASNKFKRMLNRELTHLSEMSRSGNQVSEYISNTFLDKQNDVEIPS
PTQKDREKKKKQQLMTQISGVKKLMHSSSLNNTSISRFGVNTENEDHLAKELEDLNKWG
LNIFNVAGYSHNRPLTCIMYAIFQERDLLKTFRISSDTFITYMMTLEDHYHSDVAYHNS
LHAADVAQSTHVLLSTPALDAVFTDLEILAAIFAAAIHDVDHPGVSNQFLINTNSELAL
MYNDESVLENHHLAVGFKLLQEEHCDIFMNLTKKQRQTLRKMVIDMVLATDMSKHMSLL
ADLKTMVETKKVTSSGVLLLDNYTDRIQVLRNMVHCADLSNPTKSLELYRQWTDRIMEE
FFQQGDKERERGMEISPMCDKHTASVEKSQVGFIDYIVHPLWETWADLVQPDAQDILDT
LEDNRNWYQSMIPQSPSPPLDEQNRDCQGLMEKFQFELTLDEEDSEGPEKEGEGHSYFS
STKTLCVIDPENRDSLGETDIDIATEDKSPVDT

Figure 31

a)



b)

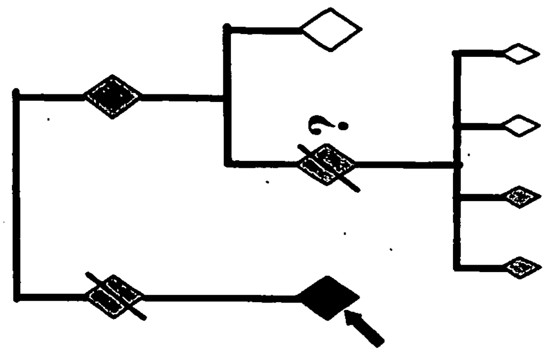


Figure 32

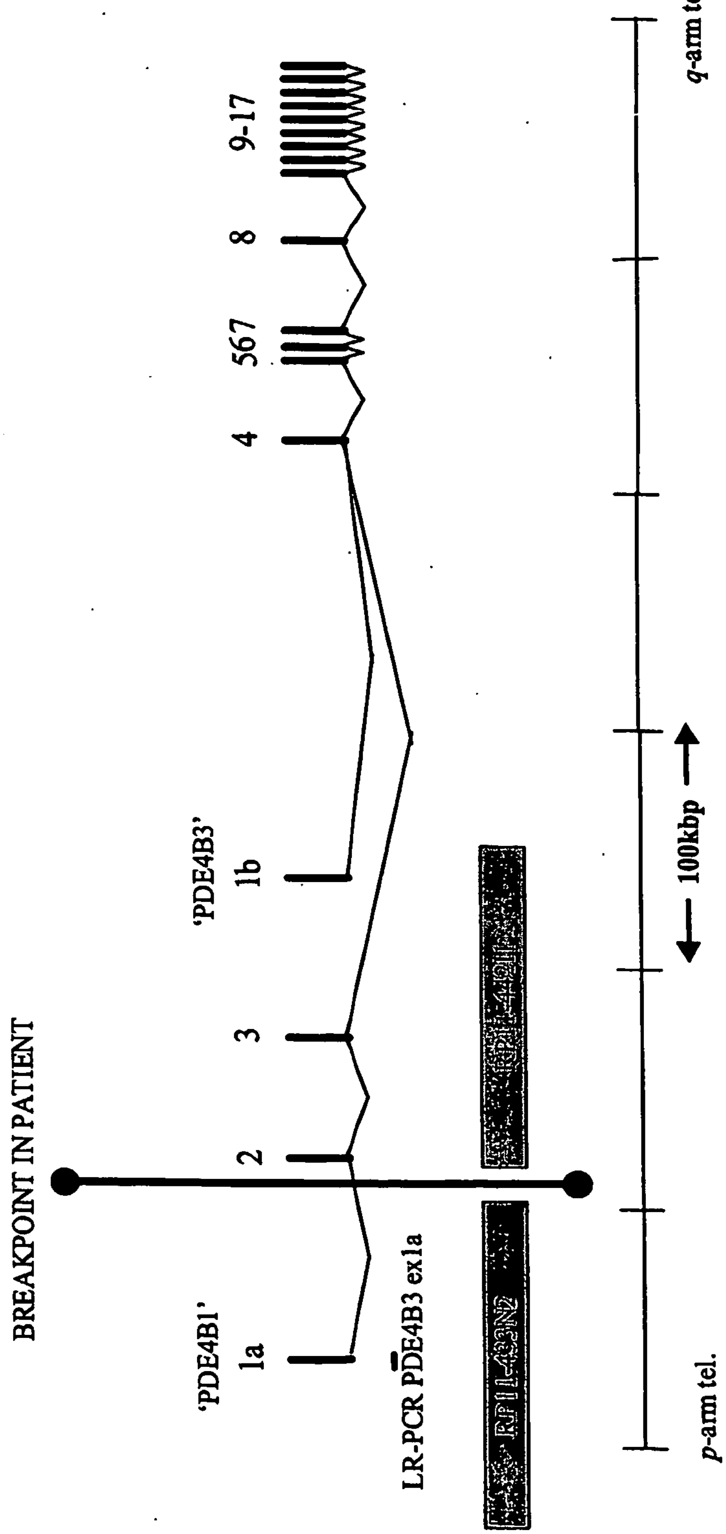

BREAKPOINT IN PATIENT
'PDE4B1'
1a
2
3
'PDE4B3'
1b
4
567
8
9-17
LR-PCR PDE4B3 ex1a
RP11-423N2
100kbp
q-arm tel.
p-arm tel.

Figure 33

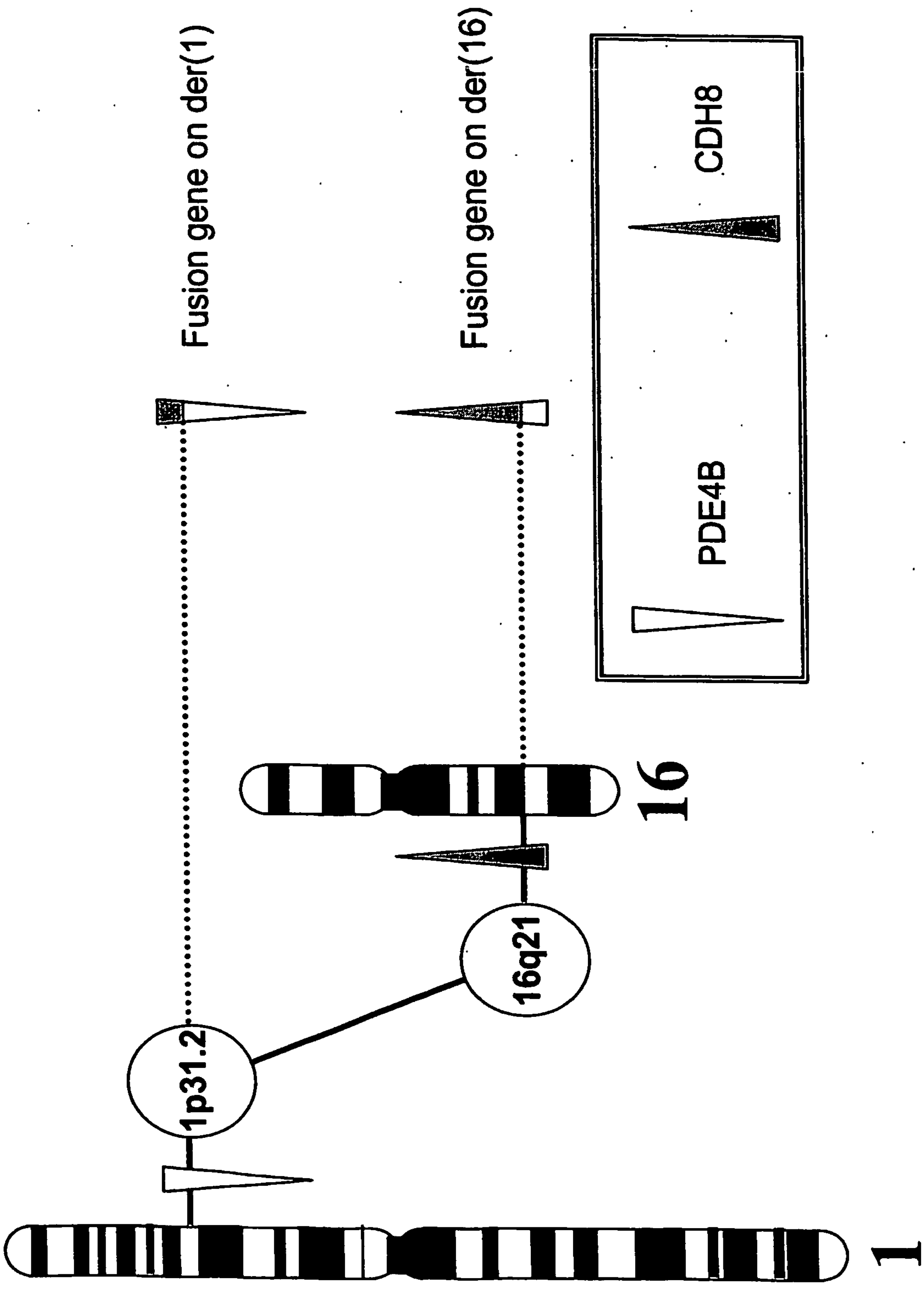

Fusion gene on der(1)
Fusion gene on der(16)
CDH8
PDE4B
16
16q21
1p31.2
1

Figure 34

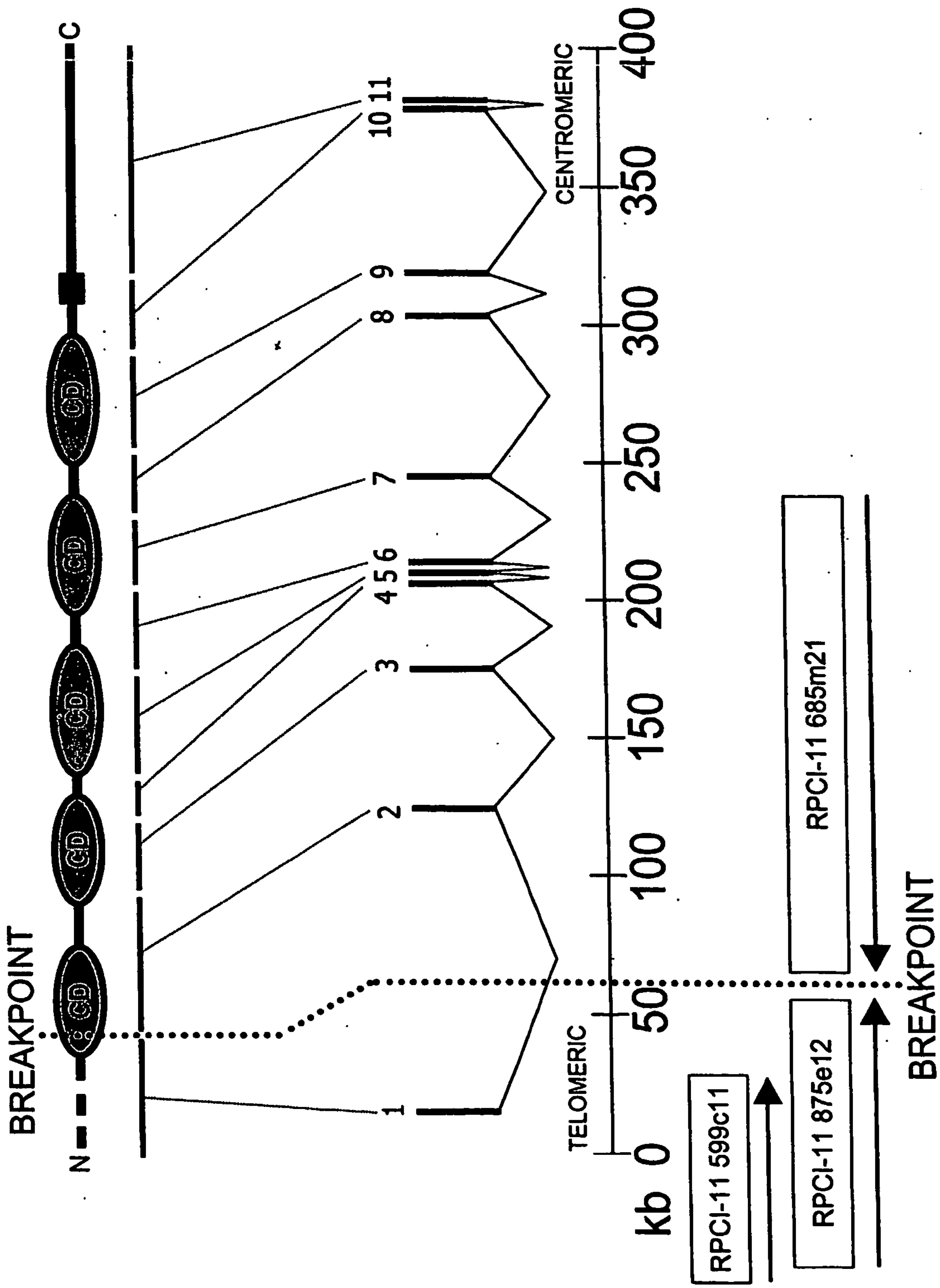

BREAKPOINT
N
C
CD
CD
CD
CD
CD
1
2
3
4 5 6
7
8
9
10 11
TELOMERIC
CENTROMERIC
kb 0
50
100
150
200
250
300
350
400
RPCI-11 599c11
RPCI-11 875e12
RPCI-11 685m21
BREAKPOINT

Figure 35

```
   1 agccatttgt gaacctggag gcttgacatt cgccagcgca gggccccaca agagaaattt
  61 caatgaaaag aaaagccaat ggattgtggt cttagaaaag ctgcttagat gatgtctgtt
 121 tcccgtgcta tagacacgtg gcagagctgt aagtaaatgc tcggcactgc atgatgaatt
 181 ggatggctgc agaccggaga caaaaaaaat aattgtctca ttttcgtggt gatttgctta
 241 actggtggga ccatgccaga acggctagcg gaaatgctct tggatctctg gactccatta
 301 ataatattat ggattactct tccccttgc atttacatgg ctccgatgaa tcagtctcaa
 361 gttttaatga gtggatcccc tttggaacta aacagtctgg gtgaagaaca gcgaattttg
 421 aaccgctcca aaagaggctg ggtttggaat caaatgtttg tcctggaaga gttttctgga
 481 cctgaaccga ttcttgttgg ccggctacac acagacctgg atcctgggag caaaaaaatc
 541 aagtatatcc tatcaggtga tggagctggg accatatttc aaataaatga tgtaactgga
 601 gatatccatg ctataaaaag acttgaccgg gaggaaaagg ctgagtatac cctaacagct
 661 caagcagtgg actgggagac aagcaaacct ctggagcctc cttctgaatt tattattaaa
 721 gttcaagaca tcaatgacaa tgcaccagag tttcttaatg gaccctatca tgctactgtg
 781 ccagaaatgt ccattttggg tacatctgtc actaacgtca ctgcgaccga cgctgatgac
 841 ccagtttatg gaaacagtgc aaagttggtt tatagtatat tggaagggca gccttatttt
 901 tccattgagc ctgaaacagc tattataaaa actgcccttc ccaacatgga cagagaagcc
 961 aaggaggagt acctggttgt tatccaagcc aaagatatgg gtggacactc tggtggcctg
1021 tctgggacca cgacacttac agtgactctt actgatgtta atgacaatcc tccaaaattt
1081 gcacagagcc tgtatcactt ctcagtaccg gaagatgtgg ttcttggcac tgcaatagga
1141 agggtgaagg ccaatgatca ggatattggt gaaaatgcac agtcatcata tgatatcatc
1201 gatggagatg gaacagcact tttgaaatc acttctgatg cccaggccca ggatggcatt
1261 ataaggctaa gaaaacctct ggactttgag accaaaaaat cctatacgct aaaggtagag
1321 gcagccaatg tccatattga cccacgcttc agtggcaggg ggccctttaa agacacggcg
1381 acagtcaaaa tcgtggttga agatgctgat gagcctccgg tcttctcttc accgacttac
1441 ctacttgaag ttcatgaaaa tgctgctcta aactccgtga ttgggcaagt gactgctcgt
1501 gaccctgata tcacttccag tcctataagg ttttccatcg accggcacac tgacctggag
1561 aggcagttca acattaatgc agacgatggg aagataacgc tggcaacacc acttgacaga
1621 gaattaagtg tatggcacaa cataacaatc attgctactg aaattaggaa ccacagtcag
1681 atatcacgag tacctgttgc tattaaagtg ctggatgtca atgacaacgc ccctgaattc
1741 gcatccgaat atgaggcatt tttatgtgaa aatggaaaac ccggccaagt cattcaaact
1801 gttagcgcca tggacaaaga tgatcccaaa aacggacatt atttcttata cagtctcctt
1861 ccagaaatgg tcaacaatcc gaatttcacc atcaagaaaa atgaagataa ttccctcagt
1921 attttggcaa agcataatgg attcaaccgc cagaagcaag aagtctatct tttaccaatc
1981 ataatcagtg atagtggaaa tcctccactg agcagcacta gcaccttgac aatcagggtc
2041 tgtggctgca gcaatgacgg tgtcgtccag tcttgcaatg tcgaagctta tgtccttcca
2101 attggactca gtatgggcgc cttaattgcc atattagcat gcatcatttt gctgttagtc
2161 atcgtggtgc tgtttgtaac tctacggcgg cataaaaatg aaccattaat tatcaaagat
2221 gatgaagacg ttcgagaaaa catcattcgc tacgatgatg aaggaggagg ggaggaggac
2281 acagaggctt ttgacattgc aactttacaa aatccagatg gaattaatgg atttttaccc
2341 cgtaaggata ttaaaccaga tttgcagttt atgccaaggc aagggcttgc tccagttcca
2401 aatggtgttg atgtcgatga atttataaat gtaaggctgc atgaggcaga taatgatccc
2461 acggccccgc catatgactc cattcagata tatggctatg aaggccgagg gtcagtggct
2521 ggctccctca gctccttgga gtccaccaca tcagactcag accagaattt tgactacctc
2581 agtgactggg gtccccgctt taagagactg ggcgaactct actctgttgg tgaaagtgac
2641 aaagaaactt gacagtggat tataaataaa tcactggaac tgagcattct gtaatattct
2701 agggtcactc cccttagata caaccaatgt ggctatttgt tttagaggca agtttagcac
2761 cagtcatcta taaactcaac cacattttaa tgttgaacca aaaaaagata ataaaataaa
2821 aaagtatatg ttaggaggtt ataaatcttg tggagtgtga attaagtatg tggagtgtct
2881 agaagtcctt ggatatttga tatttacctg accaccacag acaaagatt
```

Figure 36

1 *MPERLAEMLL DLWTPLIILW ITLPPCIYM*A PMNQSQVLMS GSPLELNSLG EEQRILNRSK
61 RGWVWNQMFV LEEFSGPEPI LVGRLHTDLD PGSKKIKYIL SGDGAGTIFQ INDVTGDIHA
121 IKRLDREEKA EYTLTAQAVD WETSKPLEPP SEFIIKVQDI NDNAPEFLNG PYHATVPEMS
181 ILGTSVTNVT ATDADDPVYG NSAKLVYSIL EGQPYFSIEP ETAIIKTALP NMDREAKEEY
241 LVVIQAKDMG GHSGGLSGTT TLTVTLTDVN DNPPKFAQSL YHFSVPEDVV LGTAIGRVKA
301 NDQDIGENAQ SSYDIIDGDG TALFEITSDA QAQDGIIRLR KPLDFETKKS YTLKVEAANV
361 HIDPRFSGRG PFKDTATVKI VVEDADEPPV FSSPTYLLEV HENAALNSVI GQVTARDPDI
421 TSSPIRFSID RHTDLERQFN INADDGKITL ATPLDRELSV WHNITIIATE IRNHSQISRV
481 PVAIKVLDVN DNAPEFASEY EAFLCENGKP GQ[redacted]IQTVSAM DKDDPKNGHY FLYSLLPEMV
541 NNPNFTIKKN EDNSLSILAK HNGFNRQKQE VYLLPIIISD SGNPPLSSTS TLTIRVCGCS
601 NDGVVQSCNV EAYVLPIGLS MGALIAILAC IILLLVIVVL FVTLRRHKNE PLIIKDDEDV
661 RENIIRYDDE GGGEEDTEAF DIATLQNPDG INGFLPRKDI KPDLQFMPRQ GLAPVPNGVD
721 VDEFINVRLH EADNDPTAPP YDSIQIYGYE GRGSVAGSLS SLESTTSDSD QNFDYLSDWG
781 PRFKRLGELY SVGESDKET

Figure 37

a)

MPERLAEMLLDLWTPLIILWITLPPCIYMAPMNQSQVLMSGSPLELNSLGEEQRILNRS
KRGWVWNQMFVLEEFSGPEPILVGRVLKSVSKLH*

b)

```
     G R G G A A E A P R A G G G R L L R G Q
3    ggccgcggcggtgcagcagaggcgcctcgggcaggaggagggcggcttctgcgagggcag   62
     P E L H T D L D P G S K K I K Y I L S G
63   cctgagctacacacagacctggatcctgggagcaaaaaaatcaagtatatcctatcaggt  122
     D G A G T I F Q I N D V T G D I H A I K
123  gatggagctgggaccatatttcaaataaatgatgtaactggagatatccatgctataaaa  182
     R L D R E E K A E Y T L T A Q A V D W E
183  agacttgaccgggaggaaaaggctgagtataccctaacagctcaagcagtggactgggag  242
     T S K P L E P P S E F I I K V Q D I N D
243  acaagcaaacctctggagcctccttctgaatttattattaaagttcaagacatcaatgac  302
     N A P E F L N G P Y H A T V P E M S I L
303  aatgcaccagagtttcttaatggaccctatcatgctactgtgccagaaatgtccattttg  362
     G T S V T N V T A T D A D D P V Y G N S
363  ggtacatctgtcactaacgtcactgcgaccgacgctgatgacccagtttatggaaacagt  422
     A K L V Y S I L E G Q P Y F S I E P E T
423  gcaaagttggtttatagtatattggaagggcagccttattttccattgagcctgaaaca  482
     A I I K T A L P N M D R E A K E E Y L V
483  gctattataaaaactgcccttcccaacatggacagagaagccaaggaggagtacctggtt  542
     V I Q A K D M G G H S G G L S G T T T L
543  gttatccaagccaaagatatgggtggacactctggtggcctgtctgggaccacgacactt  602
     T V T L T D V N D N P P K F A Q S L Y H
603  acagtgactcttactgatgttaatgacaatcctccaaaatttgcacagagcctgtatcac  662
     F S V P E D V V L G T A I G R V K A N D
663  ttctcagtaccggaagatgtggttcttggcactgcaataggaagggtgaaggccaatgat  722
     Q D I G E N A Q S S Y D I I D G D G T A
723  caggatattggtgaaaatgcacagtcatcatatgatatcatcgatggagatggaacagca  782
     L F E I T S D A Q A Q D G I I R L R K P
783  ctttttgaaatcacttctgatgcccaggcccaggatggcattataaggctaagaaaacct  842
     L D F E T K K S Y T L K V E A A N V H I
843  ctggactttgagaccaaaaaatcctatacgctaaaggtagaggcagccaatgtccatatt  902
     D P R F S G R G P F K D T A T V K I V V
903  gacccacgcttcagtggcaggggggccctttaaagacacggcgacagtcaaaatcgtggtt  962
     E D A D E P P V F S S P T Y L L E V H E
963  gaagatgctgatgagcctccggtcttctcttcaccgacttacctacttgaagttcatgaa 1022
     N A A L N S V I G Q V T A R
1023 aatgctgctctaaactccgtgattgggcaagtgactgctcgt etc..............
```